MO001/3/2015

Epidemiology

BMI2602

Semesters 1 & 2

Department of Life and Consumer Sciences

IMPORTANT INFORMATION:

This tutorial letter contains important information about your module.

BAR CODE

Learn without limits.

UNISA university of south africa

Contents

Epidemiology (BMI2602)

WELCOME MESSAGE

Welcome to Epidemiology, which forms part of Biomedical Sciences. We hope this course will broaden your understanding of epidemiology and help you further your studies.

0.1 Welcome note

Welcome to Epidemiology

Welcome to this module on epidemiology (BMI2602). Epidemiology is the study of the origin and causes of diseases in populations of people and how these diseases impact the population at large. The information generated by epidemiologists can be used to reduce the impact of disease on these population groups. The diseases investigated may have always been present in the population being studied or they may be new outbreaks.

Epidemiological studies therefore have an impact on individual patients, the community and the population overall. This is because the information can be used to improve the ability of health services to provide the best possible medical care to people and reduce the burden of disease in communities.

This module is a 12-credit course, meaning you will need at least 120 study hours for it. You will need to prepare a portfolio of your studies as you work through the material in the course. The myUnisa website will be the main teaching medium for this module. You need to visit this site frequently to interact with your fellow students and to participate in discussions about certain topics within the course. Try to go online at least once a week. We will explain the structure of the course in more detail in the next sections.

To assist you in your Epidemiology studies, we recommend that you visit epidemiology websites on the internet so that you can add to the notes that you will make from your study guide and textbook.

We hope that you will find this course thought-provoking and we wish you success on your journey.

0.2 Course information

0.2.1 How this course is organised

Before starting on the first learning unit of this study guide, we are going to present some details of the course. The course is divided into eight learning units:

- Unit 1: Introduction and dynamics of disease transmission
- Unit 2: Assessing the occurrence of disease: morbidity and mortality
- Unit 3: Diagnostic and screening tests
- Unit 4: Ways of expressing prognosis
- Unit 5: Randomised trials and drug testing
- Unit 6: Identifying the cause of a disease
- Unit 7: Causal inferences: bias, confounding and interaction
- Unit 8: Roles of genetic and environmental factors in disease causation

0.2.2 Textbook

The prescribed textbook for Epidemiology (BMI2602) that you need to use in conjunction with the online material is:

Gordis, L. 2014. *Epidemiology.* 5th edition. Philadelphia: Saunders Elsevier. (ISBN: 978-1-4557-3733-8)

The textbook is a comprehensive guide to epidemiology. You will not be required to learn the whole textbook, just the recommended reading sections. Use the online study material to guide you in what you need to learn. If you find a topic particularly interesting, then feel free to do further reading on that topic.

Note. In the study guide we will refer to the textbook as Gordis or just "the textbook".

0.2.3 Purpose of this module

This course introduces the basic concepts of epidemiology, and will emphasise measures of disease, diagnostic and screening tests and the design and interpretation of epidemiological studies. The epidemiological studies will include case-control, cohort and clinical trials.

After completing the course, you should be able to identify and apply practices, processes and principles of epidemiology in order to analyse epidemiological data to understand the origins and impact of specific diseases on populations.

We hope that you will find this course an interesting component in your studies in Biomedical Sciences.

0.2.4 Outcomes of the module

This course has been designed to give you the ability to

- define, describe and apply concepts of epidemiology to understand disease transmission and occurrence in populations and what measures are used to record the incidence of disease
- explain the limitations of diagnostic tests and what features of diagnostic tests make them valuable
- recognise different data sources and study designs that can be applied to studying the impact of diseases on populations and how this data can be used to solve problems related to disease
- show an appreciation for the mathematics required to conduct research into epidemiology, especially with regard to recognising trends in disease patterns and interpreting results
- describe how bias, confounding and interaction can affect the interpretation of causal associations and how these factors can be accounted for when analysing epidemiological studies
- discuss how both environmental and genetic factors can influence the incidence of disease
- discuss how epidemiological studies can be used to determine the role of each of these factors in specific diseases in particular populations

0.3 Distance learning

Distance studies are unique, with particular requirements for success that should not be underestimated. Once you have received your study material, plan how you will approach and complete

this module. Draw up a reasonable study schedule that can guide you through the whole module. Take into consideration the due dates of the assignments as given in Tutorial Letter 101.

0.3.1 Independent study

A crucial phase in the process of understanding and learning the basics of epidemiology is to articulate your ideas about these principles, both orally and in writing. Only when you have tried this process for yourself will you understand the full value of this exercise.

The assessment following your studies measures an aspect of success. This module will have formative (ongoing) and summative (final) assignments and examinations mainly in the form of written work. Your reflections on your process of learning are therefore also an important part of your work. Since the focus in this module is to understand and apply the concepts of epidemiology, your assessment will focus on the imbedded abilities to be competent at doing this.

We will work through the study guide, using the guidelines in the next section ("Improved study skills"). This involves drawing up mind maps and making your own summaries of the objectives and contents of chapters. Restrict summaries to one page. Be conscientious about this: it is vital to keep consulting the textbook if you are to master its contents. Additional textbooks and articles give alternative views or provide more insight into issues under discussion and are optional additional reading.

Be focused in your studies. Build up your own study and exam preparation portfolio (with your assignments, activities, reflections, summaries, self-assessments and notes) throughout your academic and/or experiential learning. This portfolio will not be assessed by the lecturer, but you will need to prepare it so that you can complete the assignments and ultimately pass the final examination. It is also very important to use this portfolio, in combination with your assignments and subsequent feedback (tutorial letters), for your exam preparation. The advantage is that you engage personally in your learning, you set goals, you evaluate your own progress through reflective actions and you evaluate your ability to realise the learning outcomes, thus becoming a more independent and self-directed learner.

What is a portfolio? A portfolio is a folder/file in which you gather and compile additional and/or summarised information during the year as you work through the study guide. This portfolio will help you to prepare for the examination by focusing on the most important facts and issues.

Your portfolio should comprise

- answers to each activity in each learning unit
- a mind map/summary of each learning unit
- your marked assignments (or a copy you made prior to submitting your assignment)
- your reflections on each learning unit
- extra reading material taken from the internet, additional books, medical and/or scientific journals
- a new vocabulary of words or glossary of new terms in your own words

To ensure that you master this module, use the following study skills guidelines.

0.3.2 Improved study skills

It is critical that you think independently and learn to look beyond the prescribed book alone. Many additional references are included in this study guide and we encourage you to consult them. In addition, as a more advanced distance education student, it is essential that you become familiar with searching for research/scientific articles on the internet.

0.3.3 How to search for research/scientific articles

As a general source of information, the Google website has now created an additional search engine under “advanced search”, called Google Scholar. This has its own advanced search with a number of linked fields. This allows a brief description of your subject query, say in four to six words, to be linked to specific phrases. On pressing “search enter”, a variety of websites relating to the query will appear. The advantage of using this portal is that you can access most of the journal references from any internet site, other than myUnisa. Certain journals, such as *Science Direct* on the Elsevier site, can only be accessed through a tertiary academic institution such as Unisa. To access this journal, you need to do this:

1. Go to Unisa online at http://www.unisa.ac.za/
2. Click on Library at the top of the page.
3. In the maroon area at the top of the page, click on “Find e-resources”.
4. Follow the guidelines if you are a first-time user.
5. Click on the option “A-Z list of electronic resources”.
6. Various links for databases will now be on your screen. Click on any database to do a search. For Epidemiology we recommend clicking on Science Direct, Nature or SpringerLink. (Remember, to find Science Direct, go to s on top and a list of all the databases starting with s will appear; and if you want to go to Nature, click on the n on top, etc.).
7. Once you enter one of these databases, you can search for scientific articles by typing in the relevant keywords in the “search” box. Be very specific with the keywords. One word will usually result in too much information and much of it won’t be on the specific topic you are looking for.
8. You will need to do some independent searches yourself, as part of your portfolio, assignments and exam preparation. Especially since this is a distance education course, you need to supplement it with information from internet sources.

Contact Unisa Library if you have any difficulties or for assistance: +27 12 429 3206 or see the Library website for your local branch library's telephone number.

0.3.4 Skimming, scanning, study-reading strategy (SSS strategy)

There are a number of strategies that can help you study, one of them being the SSS strategy. The three techniques in the SSS strategy are **s**kimming, **s**canning and **s**tudy-reading. The strategy comprises six steps altogether. What do these steps involve?

Skimming

1. **Page through and explore**. First, read the section quickly, forming a rough idea of the contents.

Concentrate on headings and subheadings, bold and italic type, boxes, tables and illustrations, and - in the case of a chapter – introductions and summaries. The objectives of a chapter are important. (Think of how you would page through a magazine. When starting a new learning unit, scan it and concentrate on the concepts that catch your eye.)

2. **Make a cursory survey**. Ask yourself while you read: What key terms occur in this section or chapter? Stop when you identify a key term and read carefully what is said about it. Mark it in the book. What you are trying to ascertain is: **Where** is it? In other words, where is the information that you will need to discuss later?

Scanning and reflecting

3. **Scan** the section or chapter.

4. **Start a mind map** (for the whole section or chapter or for parts of it, as in starting a summary). You are looking for items and concepts while reading the information in the section or chapter in a more evaluative way. Reflect on interrelationships between concepts. The question now is: **What** is it? What is the meaning and the purpose? Visualisation is important and you will certainly start writing down key concepts. You can omit parts of the text.

5. **Deeper reflection**. Start by giving your mind map a structure. As you work through the prescribed activities of the section or chapter, keep returning to the mind map to fill in the detail. Reflect on the value and meaning of categories, concepts, motivations, variables and key terms.

Study-reading

6. **Study-read**. This follows directly from stages 2, 4 and 5 and is done carefully, thoroughly and thoughtfully. The key terms and concepts you have pinpointed have to be linked up, and here the mind map and summaries are important. (Remember to put your detailed mind map in your portfolio.) Pause while reading, consolidate what you remember and consider how new information fits in with what you already have. This will give you a good idea of the whole picture.

7. **Activity-based approach**

Whenever you get to an activity in your study guide, complete it in full on loose pages which you then insert in your portfolio, grouped together per section and learning unit. Supplement this with your own notes from your portfolio. (You don't need to submit activities or the portfolio to the lecturer, but these are essential for exam preparation.)

8. **Understanding what you read**

Take time to note new vocabulary words. Use a dictionary to understand the meaning of new words, or use Google to define a word for you. You could compile a page for each learning unit and add it to your portfolio.

0.3.5 Managing your self-paced study time

You require at least 120 study hours for this module if you are an average student (this time may vary substantially). You should therefore plan to use eight study hours per week per module, meaning you

should finish one module in 15 weeks. Remember, if you have registered for more than one module, you should plan time for each module accordingly. We advise you to keep a study schedule or diary so that you have a clear idea of the time you have available for study. This will help you to manage your studies within your available time and balance study with work and family life. In Tutorial Letter 101 and on myUnisa you will find a list of due dates for various assignments, which you should keep in your normal diary. Break the large assignments into a series of smaller tasks to complete one step at a time.

To manage your workload, your best working method is to study frequently and regularly in this subject. Establish a routine in an environment with low noise and good lighting. Reward yourself after a productive session.

0.3.6 Academic specialist guidance

If you get into difficulties, please contact the staff in the Department of Life and Consumer Sciences who are responsible for this module. A positive, encouraging attitude towards the course as a whole will keep you motivated to persevere in your studies. You will be able to contact the lecturer.

0.4 Plagiarism

Never try to pass off other people's work (or our lecture notes and tutorial matter) as your own. If you wish to use other people's words and ideas or our notes in your own answers, you must use quotation marks and acknowledge your source (use the Harvard method). If you are unsure about the correct way of acknowledging sources, contact Unisa's Library Information Desk. Students who fail to acknowledge quotations or who borrow from lecture notes and outside sources or who copy someone else's answers may be refused permission to write the examination or may be penalised in the assignment.

0.5 Assessment in the module

This course will have formative (ongoing) assessment in the form of assignments and summative (final) assessment in the form of a written examination. In addition to learning a new subject (essentially a new language), it is important for you to reflect on the subject and also on your process of learning.

0.6 In conclusion

After reading this general introduction, you should now have a better understanding of what the module involves and what the aims are for completing your studies in Epidemiology. To review, learning unit 1 will introduce epidemiology as a science and the dynamics of disease transmission. Learning unit 2 will deal with measures that can be used to determine the occurrence of disease in populations. In learning unit 3 you will look at diagnostic tests and in learning unit 4 at prognosis and how it can be expressed. Learning units 5 and 6 cover different study designs that can be used to determine the efficacy of different treatments and identify the cause of disease. Learning unit 7 will include some factors that can influence the interpretation of epidemiological results and finally in learning unit 8 the role of genetic and environmental factors in disease is discussed.

0.7 Getting started

To get to know your online environment and fellow students, we would like you to work through an activity called an ice breaker.

What is an ice breaker?

An ice breaker helps you to

- understand the technologies that will be used in the course
- get to know and connect with your fellow students

There is one ice breaker activity that you need to do: a blog entry

Ice breaker: Personal blog entry

Create your own blog entry and share whether you think you would find it interesting to investigate disease outbreaks in populations as a career.

There is no right or wrong answer to the question. Go to the Blog tool by clicking on Blogs in the left-hand tool list of this website. You can find the instructions on how to use the blog in the FAQ section in the left-hand tool list of this website.

If you like, you can add links, bullets, lists and colour by using the editing buttons. You can also go back and edit your blogs. The next time we ask you to use the blog, just click on “Add blog entry” again and create a new blog, which will appear under your name.

Learning unit 1

Introduction and dynamics of disease transmission

1.1 Introduction to learning unit 1

Epidemiology is the science that deals with the occurrence, distribution and control of health and disease in defined human populations. The role of the epidemiologist is to study when and where disease occurs and how it is transmitted within populations.

In this unit we will focus on the objectives of epidemiology and some of the approaches used. In the second part of this learning unit we will consider some of the concepts and mechanisms involved in disease transmission and how epidemiology is used to analyse disease outbreaks. To work through the learning unit, refer to **chapters 1 and 2**, pages 2–37 in Gordis.

The following websites are glossaries that define common terms used in epidemiology. It may be useful to refer to them throughout this study guide when you don't understand the terminology being used.

http://www.cdc.gov/excite/library/glossary.htm

http://depts.washington.edu/physdx/eglossary.html

1.2 Learning outcomes

After completing this learning unit, you should be able to

- state the objectives of epidemiology
- explain how health problems in populations change over time
- describe different levels of prevention of disease
- discuss the link between epidemiology and clinical practice
- discuss the different modes of disease transmission
- discuss
    - clinical and subclinical disease
    - types of exposure
    - herd immunity
    - incubation period
    - attack rate
- list and discuss
    - the important questions asked when investigating an outbreak
    - the steps taken when investigating an acute outbreak

1.3 Epidemiology

Recommended reading: chapter 1, pages 2-18, in Gordis

By now you know that epidemiology is the study of the incidence, distribution and control of health and disease in human populations. So what exactly does an epidemiologist investigate? The role of

epidemiologists is varied but there are a number of core objectives of epidemiology. Some of these are

- identifying the aetiology of disease
- determining the extent of the disease
- studying the progression of the disease
- evaluating preventive and therapeutic measures for a disease
- developing public health policy

1.3.1 Changing patterns of health problems in populations over time

The primary diseases affecting human populations change over time. Refer to figure 1-2 in Gordis showing the major causes of death in the United States in 1900 and 2009. You will notice the leading causes of death are very different in the two different eras. Therefore, the kinds of interventions and services that medical practitioners need to use may differ significantly over time. Epidemiologists track these changes to guide the development of current health policies.

The change in the cause of death over the last century presented in figure 1-2 in Gordis is typically seen in the populations of developed countries. However, it is not seen in developing countries where the causes of death are still very similar to those observed in 1900. This is due to advanced medical treatment in developed countries. Therefore the health policies that are developed and implemented in different countries or regions of countries may have to be different to meet the needs of the specific populations.

Another aspect of health that has changed significantly over time in the United States is life expectancy and this is also as a result of advanced medical treatment. Refer to figure 1-3 on page 5 of the textbook.

1.3.2 Prevention of disease in populations at risk

The evidence gained from epidemiological investigation is used to identify subgroups in a population who are at high risk for disease. They can then be targeted directly during the prevention process and specific factors and characteristics that put them at risk can be recognised. Some factors that put people at risk cannot be modified (age and sex), but others can, such as diet and lifestyle choices.

There are three types of prevention:

- primary prevention
- secondary prevention
- tertiary prevention

Refer to table 1-2 on page 6 of the textbook. Read the section on primary, secondary and tertiary prevention of disease on pages 5 and 6 in Gordis.

When attempting prevention, the entire population or just a high-risk subgroup may be targeted. An example of a population-wide preventive measure would be the vaccination of all babies against a specific disease. For other less common conditions only the people at risk might be identified and screened or treated.

1.3.3 The use of epidemiology in clinical practice

Epidemiology is crucial in directing what measures are used in clinical practice. The reason for this is the diseases that are diagnosed or the treatments that are chosen depend on population data. Population data is collected during surgeries, autopsies, clinical trials and experiences with large groups of patients who have had the same disease. This data is then used by physicians to diagnose, prognosticate and select an appropriate treatment. Many illnesses have similar symptoms and their diagnosis is made easier by understanding what diseases are prevalent in populations. This is illustrated in figures 1-4 and 1-5 in Gordis.

1.3.4 Following diseases using epidemiology (the epidemiological approach)

Identifying the cause of a disease and understanding its distribution within a population is a multistep process. The first step is usually to determine if there is an association between exposure to a factor or a characteristic of the individual and the development of the disease. If patterns and associations have been found in the first step, then the epidemiologist needs to determine if there is a possible causal relationship. Does a particular factor lead to the development of disease; does it cause the disease or increase the risk of disease developing? Correlation does not imply causation – it needs to be proved.

Refer to the study on page 8 in Gordis (figures 1-7, 1-8 and 1-9) which illustrates how epidemiological studies were conducted to assess the effects of fluoridating drinking water on tooth decay.

1.3.5 History of epidemiology

The origin of modern epidemiology dates back to the mid-1800s. Much of what we know today about epidemiology is based on three famous case studies conducted by Ignaz Semmelweis, Edward Jenner and John Snow. Read the section “From observations to preventative actions”, pages 8-16 in Gordis.

1.3.6 Activity 1.1

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

a) What are the objectives of epidemiology? Why is epidemiology important?
b) Briefly discuss the changing patterns of health problems over time.
c) What is the difference between modifiable and non-modifiable risk factors?
d) Briefly describe the three types of prevention (primary, secondary and tertiary) of diseases.
e) Why is epidemiology important to clinical practice? Have you ever been to a doctor and told them that your friend just had a specific disease, or that you went to a specific location (possibly a malaria area) and that information has been used in your diagnosis?
f) When thinking about the contribution of Semmelweis, Jenner and Snow to epidemiology, imagine and note down the social, medical and scientific environment at the time that promoted or hindered their findings and how these studies were pivotal in the development of epidemiology.

1.4 Disease transmission

Recommended reading: chapter 2, pages 19–25, in Gordis

Although some diseases are essentially genetic in origin, most develop as a result of interaction between genetic and environmental factors. A number of factors can cause illness in humans including biological, chemical and physical factors, along with other less easily defined factors such as stress. Infectious or communicable diseases are diseases that can be transmitted from one person to another by direct or indirect contact. They can be caused by viruses, bacteria, parasites or fungi. Non-infectious diseases are non-transmissible; an example is cancer.

Disease occurs as a result of the interaction of the host, an agent and the environment. Figure 2-1 in Gordis shows the epidemiological triad of disease. For a host to become ill, they must be susceptible to the disease. A number of factors including the immune system, genetic susceptibility and nutritional status determine the susceptibility of a host to illness. Refer to table 2-1 in Gordis which outlines various factors that may be associated with increased risk of human disease.

Another crucial element in the infectious process is the spread of infection. Illness is not only spread from obviously sick individuals. In some diseases, certain individuals (carriers) are infected and transmit disease, yet remain healthy; some diseases are contagious during incubation, and still other diseases are acquired from animal or environmental sources.

1.4.1 Modes of disease transmission

Diseases are transmitted either directly or indirectly.

Direct transmission from one person to another can occur by

- droplet transmission (coughing or sneezing on another person)
- direct physical contact (including sexual contact)
- transplacental transmission (mother to child)

Indirect transmission from one person to another can occur by

- indirect physical contact (touching contaminated surfaces)
- airborne, food-borne and waterborne transmission (generally through contaminated food or water)
- a vector such as an insect or an intermediate animal host

The rate and magnitude of the spread of infectious organisms through populations depends on its method of transmission and on the infectious agent's growth characteristics. The skin and mucous membranes are the primary defence against infectious agents. To cause disease microorganisms need to pass these barriers. The most common portals of entry are

- respiratory tract (upper and lower airways)
- gastrointestinal tract (mouth)
- genital area
- urinary tract
- damaged skin (cuts, burns etc.)

Refer to figure 2-3 on page 21 of the textbook.

Once someone becomes infected, they may or may not show disease symptoms. The majority of infections do not cause symptoms and are thus classified as non-clinical disease. Only a small portion of infections result in clinical disease. This is called the iceberg concept of infectious diseases and is shown in figure 2-4 on page 21 of the textbook.

Non-clinical disease can be preclinical, subclinical, persistent or latent. These terms are defined on page 22 of Gordis. A carrier is an individual that harbours the infectious organism, but shows no symptoms and serological studies are negative. The concern is that this individual can still infect others but may be unaware that they are infective. A well-known case was that of Typhoid Mary who was a carrier of *Salmonella typhi* and infected many people over several years.

1.4.2 Terminology - endemics, epidemics and pandemics

It is important for any epidemiologist to be familiar with certain epidemiological terminology. Three of the more frequently used terms are endemic, epidemic and pandemic. See the definitions on page 23 of the textbook and read the section “Endemic, epidemic and pandemic” in Gordis.

1.4.3 Activity 1.2

Do the following activity and add it to your portfolio:

a) Define the following terms:
   - carrier
   - endemic
   - epidemic
   - pandemic
   - susceptibility

b) Discuss how disease can be transmitted.

c) Describe the difference between clinical and subclinical disease.

1.4 Disease outbreaks

Recommended reading: chapter 2, pages 25–36, in Gordis

There are different types of exposures that can lead to a disease outbreak. Where a group of people have eaten contaminated food and become ill, this is referred to as common-vehicle exposure. It may be that they are only exposed once, called a single exposure, or they may be exposed more than once, called multiple exposure. It is also important to determine whether the exposure is periodic or continuous. The extent of disease that results in a population depends on the level of susceptibility within the population. Susceptibility will depend on the immunity and genetic factors of the population affected. If a larger portion of the population is susceptible, then the magnitude of the outbreak will be greater. Refer to Gordis page 25, which describes a single common-vehicle outbreak.

1.5.1 Herd immunity

Herd immunity is based on the principle that if a large percentage of the population is immune to a particular disease, the entire population is protected, not just the individuals who are immune. This occurs because if someone becomes ill and most people are immune, the ill person is unlikely to come into contact with another susceptible individual and therefore the disease does not spread.

This is especially important when viewing immunisation programmes where not every individual gets immunised, yet if a large enough percentage of the population is immunised, then the disease will not spread. Refer to figure 2-11 on page 27 of the textbook which illustrates the effect of herd immunity.

1.5.2 Incubation period

The incubation period is the time it takes for a disease to develop after exposure to a pathogenic organism or another agent (chemical or radiation). Disease may take days or weeks to develop. During this incubation period, a person may feel completely well and show no signs of disease.

The incubation periods of diseases can vary. The problem is that during at least part of the incubation period, the disease can be transmitted to other susceptible individuals. Therefore, to stop the spread of a disease, any individuals that have come into contact with infected individuals may need to be isolated for a time; this is called quarantine. It does not help to isolate only clinically ill people as they may already have transmitted the disease before they showed symptoms. Refer to figure 2-12 in the textbook which shows the incubation periods of a number of viral diseases.

In a single-exposure, common-vehicle epidemic, the epidemic curve is a distribution of incubation periods (refer to figures 2-13 and 2-14) and has a characteristic configuration.

1.5.3 Predicting the number of cases in an epidemic

There are three important questions to answer when investigating an outbreak:

1. When did the exposure take place?
2. When were symptoms apparent?
3. What is the incubation period?

Attack rate can be used to predict the number of cases expected during an epidemic and is used when analysing disease outbreaks. It is defined as the number of new cases (people who develop disease) in the population at risk divided by the total number of people at risk in the population. Therefore if you know the attack rate and the number of people at risk, you can predict how many people you expect to become ill during an outbreak. An example of attack rates can be seen in table 2-5.

People who become ill may be defined as primary or secondary cases depending on the source of their exposure. Secondary cases acquire disease from primary cases and secondary attack rates can also be calculated.

1.5.4 Exploring the occurrence of disease

When investigating the occurrence of a disease, there are three critical questions the epidemiologist must ask:

1. Who was attacked by the disease (sex, age, race)?
2. When did the disease occur (winter, summer, year round)?
3. Where did the cases arise (across the entire country or in an isolated region)?

Refer to figures 2-15 to 2-21 and the associated text, which illustrate different distributions of disease with regard to characteristics of the cases, periodicity of disease and geographical distribution.

1.5.5 Outbreak investigation

There are some common steps that are followed in an outbreak investigation. These are outlined in table 2-4 on page 35 of the textbook. In practice, the exact order of the steps often varies and several steps may be taken at the same time.

When there are several possible causal agents, cross-tabulation is a helpful method for determining which of the possible agents are responsible. The use of cross-tabulation in a food-borne outbreak of an infectious disease is illustrated in table 2-6 in Gordis.

1.5.6 Activity 1.3

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

a) Explain each of the following terms:
   - common-vehicle exposure
   - single exposure
   - multiple exposures
   - periodic contamination
   - continuous contamination

b) What factors of a disease affect the incubation period?

c) Briefly discuss
   - herd immunity
   - incubation period

d) What is meant by attack rate?

e) What is the difference between primary and secondary attack rates?

f) What critical questions are asked during a disease outbreak?

g) List the steps to be taken during a disease outbreak.

h) When is cross-tabulation used?

i) Complete the review questions at the end of chapter 2 in Gordis.

In addition, the following are some articles on the work covered in this learning unit. It is not essential

that you read them, although you may find them interesting:

http://ije.oxfordjournals.org/content/24/4/655.full.pdf

http://www.banffpork.ca/proc/2005pdf/BO09-HardingJ.pdf

1.5.7 Activity 1.4

Discussion forum question

This is a discussion activity that you should answer in the Discussions tool on the module website. If you don't have internet access, do the activity in writing and add it to your portfolio.

Do an internet search, or consult other relevant resources, and identify a disease outbreak that has recently been reported in South Africa.

The article at http://www.sajei.co.za/index.php/SAJEI/article/download/388/503 may be of assistance or the NICD web page (http://www.nicd.ac.za/) to get you started. Also try Googling "disease outbreak South Africa".

Then go to the discussion topic entitled: South African disease outbreaks, and post a summary of your findings on one outbreak you found interesting (± 250 words).

Also read the other students' postings and respond to at least one other posting. Mention anything you found particularly interesting about the other posting, or ask for clarification or more details of any aspect if you wish.

Learning unit 2

Assessing the occurrence of disease: morbidity and mortality

2.1 Introduction to learning unit 2

For epidemiologists to be able to analyse/study the transmission of diseases in human populations, they need to be able to assess the frequency of disease (how many people have been ill) and the number of disabilities and deaths that have occurred as a direct result of the disease. Morbidity describes the rate of incidence of disease in a population of people. Mortality refers to the number of people who have died in a population.

In this study section we will focus on measures of morbidity and mortality. We will discuss how rates and proportions are used to express the extent of morbidity and mortality in a population resulting from a specific cause. To work through the learning unit, refer to **chapters 3 and 4**, pages 38–87 in Gordis.

2.2 Learning outcomes

After completing this learning unit, you should be able to

- define and calculate the various measures of morbidity, including incidence, prevalence and attack rate
- describe the relationship between incidence and prevalence
- discuss active and passive disease surveillance and its importance
- describe the different measures of mortality and how they can be calculated
- explain some of the problems with mortality data and how to overcome them
- describe artificial and real differences of trends in mortality and define the cohort effect
- define the DALY and explain its importance

2.3 Information available to measure disease

Recommended reading: chapter 3, pages 38-41, in Gordis

Information to calculate the frequency of both disease occurrence and deaths within a population from a specific cause can be gathered from a number of sources including

- hospital records
- physicians' records
- interviews with patients
- medical aid records
- death certificates

The source of data from which cases are identified will influence the rates you calculate for expressing the frequency of disease. You must therefore consider the characteristics and accuracy of the sources of the data before interpreting rates and comparing them to rates from other populations.

During the development of disease in a population, only a portion of the population will become clinically

ill and display symptoms. Of that group a fraction will seek medical care and, depending on the severity of the illness, a portion of this group may need to be hospitalised. Refer to figure 3-1 on page 40 of Gordis.

Examining figures 3-1 and 3-2, you can see how the source of data used to assess the number of cases of disease will detect different numbers of cases. The number of cases detected is therefore related to the severity of the illness and the source of the data. Each source that is used to generate statistical data has its limitations and these limitations need to be taken into account when the data is interpreted.

2.4 Morbidity

Recommended reading: chapter 3, pages 41–58 of Gordis

How do we define morbidity (or illness) in a population? Typically the occurrence of disease in a population is described using rates (how fast a disease is occurring) or proportions (what fraction of the population is affected). We will discuss some of these measures of morbidity now in greater detail.

2.4.1 Incidence rate

Incidence rate is the number of new cases of a disease that occur in a population in a given period.

$$\text{Incidence rate per } 1\ 000 = \frac{\text{Number of } \mathbf{new} \text{ cases occurring during a given period}}{\text{Number of people at risk during the same period}}\ X\ 1\ 000$$

As incidence rate only includes new cases of disease, it measures the risk of developing the disease within a specific period. The possibility of developing disease can be limited to a specific population group, for example children under 5 years of age, or only females. An important point is that any individual who is included in the denominator must have the potential to become part of the group in the numerator.

The number of people at risk during the period (denominator) can be calculated in two ways:

- All the people at risk are observed throughout the entire period and their disease status recorded (the choice of period is unimportant but must be clearly stated).
- Not all the people at risk are observed for the full period. If different people are monitored for different periods, then the denominator is calculated by taking the sum of the units of time that each person at risk was monitored. This type of measure is typically called person-time and the incidence rate is calculated according to the equation below. Refer to figure 3-4 on page 43 and figure 3-5 on pages 44–45 in the textbook which illustrate the concept of person-time.

$$\text{Incidence rate per } 1\ 000 = \frac{\text{Number of } \mathbf{new} \text{ cases occurring during a given period}}{\text{Total person time (Sum of the periods of observation of each person)}}\ X\ 1\ 000$$

2.4.2 Attack rate

Remember we discussed attack rate in the previous learning unit? Attack rate was used to aid in the study of a food-borne disease outbreak and was defined as the number of people exposed to suspect food who became ill, divided by the number of people who were exposed to that food. The period is not

explicitly defined, and so this rate is thus not truly a rate, but rather a proportional measure.

Two important points regarding attack rate are that

- time is specified implicitly
- attack rate is a proportion rather than a rate

Identifying new cases in order to calculate incidence

When we wish to calculate incidence, how do we identify all new cases in a population during a specified period?

It is not always possible to monitor an entire population over time. Instead, the identified population can be screened at baseline for the disease. Those who do not have the disease are then followed for a specified time and then rescreened (refer to figures 3-6 and 3-7 on page 46 of the textbook). Any new cases identified during rescreening developed during the specified period and can function as the numerator for the incidence rate.

Sometimes only information regarding the number of cases is assessed. See figure 3-8 in the textbook – although limited, this information may still be useful. The incidence rate is required if comparisons are to be made between two populations (see figure 3-9 in the textbook).

2.4.3 Prevalence

Prevalence is a measure of how commonly a disease occurs in a population. It is defined as the total number of cases of disease in a given population at a specified time divided by the number of individuals in the given population at a specified time. Prevalence is not a measure of risk, but it is an important and useful measure of the burden of disease in a community.

There are two types of prevalence:

- point prevalence (number of individuals with the disease at a specific point in time)
- period prevalence (how many people have had the disease at any point during a specified period)

Table 3-1 illustrates the difference between the two types of prevalence. Refer also to figures 3-10 to 3-14, and the associated text, which describe the relationship between incidence and prevalence. Figures 3-15 and 3-16 are examples of prevalence data.

2.4.4 Incidence and prevalence measurements

Problems with measurements of incidence and prevalence

There can be a number of problems when generating incidence and prevalence data.

Problems with numerators	**Problems with denominators**
• Defining who has the disease (this is especially relevant for diseases that are	• Classifying individuals into certain population groups is not always clear.

hard to diagnose, e.g. dementia and rheumatoid arthritis)
- Deciding which people should be included in the numerator owing to possible sources of error in data gathering or differences in classification of disease (refer to table 3-4 in Gordis)
- Everyone represented by the denominator must have the potential to enter the group represented by the numerator.

Epidemiological studies often rely on data from hospital records. There are, however, some problems that arise when using hospital data for research purposes. These are summarised in table 3-5 on page 54.

Relationship between incidence and prevalence

Earlier in this learning unit we mentioned the difference between incidence and prevalence: incidence is a measure of risk and prevalence is not, since it does not take into consideration the duration of the disease. There is, however, an important relationship between incidence and risk demonstrated by the following equation:

Prevalence = Incidence rate × Duration of the disease

Therefore, prevalence depends on incidence rate and duration of disease.

2.4.5 Activity 2.1

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

a) Define and explain the difference between proportion and rate.
b) What is incidence rate and how is it calculated?
c) A population of 200 000 people were monitored for a year. During this time 70 people became ill with mumps. Calculate the incidence rate per 1 000 for mumps in that year.
d) What is the difference between incidence and prevalence and how are the two measures related?
e) Briefly describe some problems that may arise when using hospital data for epidemiological studies (describe at least three points).
f) Refer to figure 3-20 in Gordis showing the breast cancer incidence rates in women by age and the distribution of breast cancer in women by age. Also discuss why less than 5% of the cases occur in the oldest age group of women even though the incidence increases dramatically with age?

2.4.6 Feedback on activity 2.1

The incidence rate of mumps for question c is 0.35 per 1 000 individuals.

The limitations of hospital data (question e) are tabulated in table 3-5 in Gordis.

2.4.7 Surveillance

Disease surveillance refers to monitoring the spread of disease by collecting and analysing data about patterns of disease. It is an important role of public health and should guide the planning, implementation and evaluation of public health systems.

Some of the information gathered during disease surveillance is about

- morbidity
- mortality
- infectious disease patterns
- non-infectious disease patterns
- completeness of vaccination coverage
- levels of environmental risk factors for disease

The data collected provides decision-makers with guidance for planning, implementing and evaluating programmes for disease prevention and control. There are two types of surveillance, namely active surveillance and passive surveillance. Refer to page 39 in Gordis which describes these two types.

Quality of life

Most diseases will affect a person's quality of life. Both lethal and non-lethal diseases can cause significant distress and disability. Finding measurements that assess quality of life appropriately is still difficult, but there is general agreement that these measures are necessary to establish prioritisation of healthcare resources.

2.4.8 Activity 2.2

Do the following activity and add it to your portfolio:

a) What is surveillance and why is it important?
b) Discuss the difference between active and passive surveillance measures.
c) Why should quality of life be taken into account, and not just morbidity and mortality data, when assessing the impact of a disease on a population?
d) Complete the review questions at the end of chapter 3 in the textbook.

2.4.9 Feedback on activity 2.2

Surveillance measures are discussed extensively on page 39 in Gordis.

Quality of life is important because certain diseases may leave people disabled where they are unable to work or look after themselves. Even though these people may no longer have the disease and did not die, they have still been greatly affected by it and may be a burden on the population. Other diseases may be chronic and non-life threatening but may cause pain or impact on a person’s ability to carry out activities of daily living.

2.5 Mortality

Recommended reading: chapter 4, pages 61–85 of Gordis

In this section you will learn about mortality and describe the methods and approaches for using mortality data in epidemiological investigations.

Mortality is of great interest because

- it can be used to indicate the severity of a disease and provide information on whether the treatment for a disease has become more or less effective over time
- it can identify differences in the risk of dying from a disease between people in different geographical areas and subgroups in the population
- mortality rates can be used instead of incidence rates when the disease being investigated is severe and lethal

Examine figures 4-1 to 4-5 in the textbook which illustrate the use of mortality data in various epidemiological studies.

2.5.1 Mortality rate

The mortality rate (or death rate) is a measure of the number of deaths in a population, scaled to the size of that population, per unit time (the period selected is arbitrary, but it must be stated). The annual mortality rate owing to all causes can be calculated by dividing the total number of deaths from all causes in one year by the number of people in the population at midyear.

$$\text{Annual mortality rate owing to all causes (per 1 000 population)} = \frac{\text{Total number of deaths from all causes in one year}}{\text{Number of people in the population at midyear}} X\ 1\ 000$$

Mortality rates can also be restricted to a certain group in a population, for instance mortality in children younger than 10 years old or mortality owing to a specific disease. When a restriction is placed on a rate, it is essential that everybody in the group represented by the denominator be able to enter the group represented by the numerator.

The crude mortality rate does not make allowance for the fact that the likelihood of dying depends on age, sex, socioeconomic class and other factors. This is important to remember when comparing mortality rates of two different populations.

Case fatality rate

Case fatality rate is the percentage of deaths within a designated population of people with a particular condition over a certain period.

$$\text{Case fatality rate (\%)} = \frac{\text{Number of people who die during a period after disease diagnosis or onset}}{\text{Number of people with the specific disease}} X\ 100$$

Case fatality rate is therefore the percentage of people who die within a certain period after being diagnosed with a certain disease. Unlike mortality rate, case fatality rate is a measure of the severity of

the disease. Refer to table 4-1 on page 65 of the textbook which outlines the difference between mortality rate and case fatality rate.

Proportionate mortality

Proportionate mortality is another measure of mortality. It is not a rate but rather a ratio. It can be defined as the number of deaths from a specific cause per 100 or 1 000 deaths from all causes in the same period. Refer to figures 4-6 and 4-7, and tables 4-2 to 4-4 on pages 66–68 in the textbook.

Proportionate mortality gives an indication of the major causes of death; it does not specify the chance of a person dying from a disease.

Years of potential life lost

Death occurring at a younger age involves a greater loss of future productive years than death occurring at an older age. This is expressed as years of potential life lost (YPLL).

It can be calculated in two steps:

1. For each cause, each deceased person's age at death is subtracted from a predetermined age at death (generally 65 years).
2. The YPLL for each individual are then added together to yield the total YPLL for a specific cause of death.

See figures 4-9 and 4-10 and table 4-5 in Gordis.

2.5.2 Why consider mortality?

Mortality is an index of the severity of a disease and can also be used as an index of the risk of disease. However, mortality is only a good measure of risk or incidence of disease when

- the case fatality rate is high
- the duration of the disease is short

When the disease is mild and not fatal, mortality is not a good measure of incidence of disease. Morbidity is then a better measure.

Problems with mortality data

Obtaining accurate mortality data can be problematic because most of our information about deaths comes from death certificates. On death certificates, deaths are coded according to underlying cause of death using the international classification of diseases (ICD). Unfortunately, the underlying cause of death does not contain information on the immediate cause of death or any other contributory causes of death. An example would be somebody with HIV dying from pneumonia; the death certificate may indicate either illness as cause of death. If an epidemiologist wishes to calculate the number of people who died of HIV and it was recorded that a specific person died of pneumonia, their death will not be included.

Also note that the ICD is reviewed and adjusted continuously. Thus differing trends in mortality can be due partly or entirely to changes in the ICD. Refer to figures 4-22, 4-23 and 4-24 and the explanatory text in Gordis, on pages 75-76.

When there is an increase or a decrease in mortality, always ask whether it is real or as a result of differences in data collection.

2.5.3 Comparing mortality in different populations

Mortality data can be used to compare two or more populations or one population at different times. Even though there are many characteristics that effect mortality, age is the biggest predictor of mortality. Therefore it is important to take into account the age distribution of two populations when they are being compared.

Table 4-7 shows a crude or unadjusted mortality rate by race. According to this table, mortality is higher in the white population than in the black population even though at the time (1965) the black population had poorer living conditions and less access to medical care. Why is this the case – is the data real? In this example, the white population is older than the black population (remember that mortality increases with old age); in the white subgroup, the greater number of deaths (from old age) increases the crude mortality rate. This will then mask the fact that the black population has a higher risk of mortality than the white population in each age group (refer to table 4-8).

There are two approaches that are used to account for age differences in two populations:

- direct age adjustment
- indirect age adjustment

In direct age adjustment, a hypothetical standard population is created. This is then used to reduce the effects of any age differences between two or more populations being compared. See tables 4-9 to 4-12 and the explanatory text in the textbook which illustrate how data can be analysed using direct age adjustment.

Age-adjusted rates are very valuable when making comparisons between populations. However, age-specific mortality data should also be examined before adjustment for any noteworthy differences or changes. Interesting changes may be masked by the age-adjusted rates, and will not be detected if only the age-adjusted rates are analysed.

In indirect age adjustment the number of expected deaths in each age group in the population of interest is calculated and added together. The number of deaths that were actually observed in the same population are likewise added together. The total number of deaths actually observed is then divided by the total number of expected deaths in the population of interest. The resultant ratio is called the standardised mortality ratio (SMR).

Refer to the example in table 4-13 in Gordis which shows how SMR values are generated.

2.5.3.1 The cohort effect

A cohort is a group of people who share the same experience. When people are born at about the same time (year or decade), they will share various characteristics as a group (a cohort), and this may affect

their susceptibility to disease. For instance, if there were a few years of famine in a region, then it may have affected the development of the children who lived through the time of famine versus children who were born 10 years later and received enough nutrients while growing up. See table 4-15 on page 83 of the textbook which illustrates the cohort effect.

It is then important that when we examine changes in mortality over time, we always take into account that there may be a cohort effect.

2.5.3.2 Interpreting observed changes in mortality

Differences in mortality over time or between populations may be artificial or real. Possible reasons for artificial results are given in table 4-16 in Gordis. Some reasons for real changes in mortality over time are given in table 4-17 page 83 of the textbook.

2.5.3.3 Projecting the future burden of disease

In an attempt to predict the future burden of disease on the world, a study called the Global Burden of Disease was carried out. The study examined mortality data and the impact of premature death and disability on a population. These factors were combined and an index was developed called the disability adjusted life year (DALY). The findings of the Global Burden of Disease study are shown in table 4-18 and figures 4-25 and 4-26.

There is currently no universal agreement on how to determine or apply a single measure of disease burden (such as the DALY). Such a universal measure, however, would be useful to enable valid comparisons between populations, informing the development and implementation of appropriate interventions worldwide.

2.5.4 Activity 2.3

Do the following activity and add it to your portfolio:

a) Define and explain how to calculate the following measures:
   - mortality rate
   - case fatality rate
   - proportionate mortality
   - years of potential life lost

b) What is the difference between case fatality rate and mortality rate?

c) Describe some of the problems that occur when generating mortality data from death certificates.

d) Why is age so important when comparing mortality in different populations and what approaches can be used to overcome the problem that arises when dealing with age?

e) In your own words describe what the cohort effect is.

f) What are some real and artificial explanations for changing trends in mortality?

g) What does DALY stand for?

h) Complete the review questions on pages 82–84 of the textbook.

2.5.5 Feedback on activity 2.3

Are you able to calculate the different measures of mortality if you are given the relevant data?

A comparison of mortality rate and case fatality rate is illustrated well in table 4-1 in Gordis.

You should have referred to the studies illustrated in figures 4-20, 4-21 and 4-22 when describing some of the problems that occur when generating mortality data from death certificates.

Possible reasons for artificial results are listed in table 4-16 and reasons for real changes in mortality are given in table 4-17 of the textbook.

Learning unit 3

Diagnostic and screening tests

3.1 Introduction to learning unit 3

When studying disease transmission and aetiology, it is important to be able to differentiate between people who have the disease and those who do not. Screening and diagnostic tests are used to determine which people in a population have the disease and which do not. Screening is used to identify individuals in a population that have the disease but may not know they have it. If individuals with the disease are identified, it is hoped that the mortality and suffering from the disease can be reduced.

In this learning unit we will discuss how the quality of newly available screening and diagnostic tests is assessed, and how they can be used in prevention programmes. To work through the learning unit, refer to **chapter 5**, pages 88–113 in Gordis.

3.2 Learning outcomes

After completing this learning unit, you should be able to

- define the following terms: validity, sensitivity, specificity, true positive, false positive, true negative and false negative
- explain the difference between a dichotomous test and a continuous variables test
- explain the consequences of false positive and false negative test results
- explain the use of multiple diagnostic tests
- calculate sensitivity, specificity, positive and negative predictive value and overall percentage agreement
- explain the effect of prevalence and specificity on predictive value of a test
- discuss the validity and reliability of diagnostic tests

3.3 Biological variation of human populations

Recommended reading: chapter 5, pages 88-89, in Gordis

People are biologically different and so when using diagnostic tests to distinguish between individuals with a condition and those without, it is important to understand that people have different characteristics and may respond differently to screening tests. Refer to figures 5-1 and 5-2 which illustrate distributions of results obtained in screening tests.

3.4 Validity of screening tests

Recommended reading: chapter 5, pages 89-95, in Gordis

A screening test is valid if it is able to distinguish between people who have a disease and those who do not. The test needs to be sensitive enough to be able to identify the people who have the disease but also specific enough that it does not identify people as having the disease when in actual fact they do not have it.

Screening tests can be divided into two groups:

- tests with dichotomous results
- tests of continuous variables

Tests with dichotomous results

This type of test only has one of two possible results – positive or negative. It is important to know how good the test is at identifying diseased individuals. This is done by calculating the sensitivity and specificity of the test. Refer to table 5-1 on page 90 of the textbook and make sure you can calculate the sensitivity and specificity of a screening test. If a test does not correctly identify the disease status of people, this is referred to as a false positive or false negative. Refer to table 5-2 in Gordis.

Tests of continuous variables

These types of tests examine a continuous variable such as blood pressure. There is no positive or negative result. Rather, a cut-off level is established above which a test result is considered positive and below which a result is considered negative. Refer to figure 5-3 on page 92 of the textbook.

It is essential that the correct cut-off level be established otherwise people will be incorrectly classified as having the disease when they do not, or as not having the disease when they do. There is no perfect cut-off level; the objective is to determine the most beneficial cut-off level. If the cut-off level is too high or too low, then false positives and false negatives will result (refer to figures 5-5 and 5-6 in Gordis). It may be more beneficial to have more false positives that can be screened further, but this is expensive. If the disease is serious and there are too many false negatives, those people will not get the treatment they need. The choice of cut-off level then depends on the relative importance of false positivity and false negativity for the disease in question.

3.4.1 Activity 3.1

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

a) Define the terms “validity”, “sensitivity” and “specificity” with regard to screening tests.
b) Explain the difference between tests with dichotomous results and tests of continuous variables.
c) Why is it important that a screening test be accurate? Be sure to mention true positive, false positive, true negative and false negative test results and the implications of getting false negative and false positive results.

3.5 Use of multiple screening tests

Recommended reading: chapter 5, pages 95-100, in Gordis

Different screening tests may be used either sequentially or simultaneously to test certain individuals.

Sequential (two-stage) testing involves the use of two tests: the first test is performed and, depending on the results, a second test may be performed. The less expensive, less invasive, less uncomfortable

test is generally performed first. The second test will typically be more sensitive and specific. See the example in figure 5-7, page 95 in Gordis. When using sequential tests the net specificity and sensitivity need to be determined. Refer to the calculation for this example on pages 95-96 in Gordis. Make sure you can calculate the net specificity and sensitivity of sequential testing.

In simultaneous testing two tests are done at the same time and the results are used to calculate the net sensitivity and net specificity. Refer to tables 5-3 to 5-6 and figures 5-8 and 5-9 which show how the net sensitivity and net specificity can be calculated for two simultaneous tests.

Comparison of simultaneous and sequential testing

When deciding whether to use simultaneous or sequential testing, the following must be taken into consideration: When two sequential tests are used, those who test positive on the first test are brought in for the second test. There is a loss in net sensitivity but a net gain in specificity compared with either test done alone. When two simultaneous tests are used, there is a net gain in sensitivity and a net loss in specificity compared with either test alone.

The choice of which type of testing to perform will depend on whether the test is done for screening or diagnostic purposes and on practical considerations, for example cost and degree of invasiveness of the tests being performed.

3.5.1 Activity 3.2

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

> Explain what sequential testing and simultaneous testing are. How do the different types of testing procedures affect the net specificity and net sensitivity of the screening process?

3.6 Predictive value of a test

Recommended reading: chapter 5, pages 100-105, in Gordis

An important question to ask in public health is what percentage of people who have a disease will be correctly identified by a specific diagnostic test? This will, of course, come down to how accurate the test is at identifying people who are ill and people who are not. This leads to two questions: What proportion of people who tested positive actually have the disease? What proportion of people who tested negative actually do not have the disease? To answer these questions, we can calculate positive predictive value or negative predictive value. Refer to table 5-7 on page 100 of the textbook, which shows how to calculate the positive predictive value and negative predictive value of a test. Make sure you are able to calculate these values.

The sensitivity and specificity of a test are characteristics of the test being used. However, the predictive value is affected by two factors:

- the prevalence of the disease in the population being tested
- the specificity of the test being used when the disease does not occur frequently

Relationship between positive predictive value and disease prevalence

There is an important relationship between positive predictive value and disease prevalence: the higher the prevalence, the higher the positive predictive value. This is shown in table 5-8 on page 101, and figure 5-11 on page 102 of the textbook.

This is important because it means that a screening programme is most beneficial and economical if it is directed at a high-risk target population. Screening populations where the disease occurs infrequently will yield only a few undetected cases for the amount of screening that is conducted. In high-risk populations screening will yield more cases for the amount of screening and will therefore be more economical.

Because of the relationship between predictive value and disease prevalence, the results obtained need to be analysed while keeping in mind the prevalence of disease in the population being tested. This is illustrated in the example shown in figure 5-12 and table 5-9 in the textbook.

Relationship between predictive value and specificity of the test

An increase in specificity of a test will result in an increase in the positive predictive value. This is illustrated in figure 5-13 and table 5-10 on page 101 of the textbook.

3.6.1 Activity 3.3

Do the following activity and add it to your portfolio:

a) Define positive predictive value and negative predictive value.
b) What is the relationship between the following?
    - positive predictive value and disease prevalence
    - positive predictive value and specificity

3.7 Reliability of tests

Recommended reading: chapter 5, pages 105-110, in Gordis

If test results are not reliable and cannot be reproduced, the value and usefulness of the test are minimal. There are a number of factors that contribute to variation between test results:

- **Intrasubject variation (variation within individual subjects):** There is great variation of many human characteristics over time. Table 5-11 in Gordis shows the variation in blood pressure readings during a 24-hour period for three individuals. Since different conditions can lead to different results, the conditions under which the tests are performed must be kept as similar as possible.
- **Intraobserver variation (variation in the reading of test results by the same reader):** The same reader who looks at the same test results more than once may come to a different conclusion the second time.
- **Interobserver variation (variation in the reading of test results by different readers):** Two observers looking at the same test often do not draw exactly the same conclusion from the results. The extent to which they agree or disagree is so important that we need to be able to

express it quantitatively. This can be done by calculating

- the overall percentage agreement
- the kappa statistic

Refer to table 5-12 and figure 5-15 for the method used to calculate the overall percentage agreement between two observers. You do not need to be able to determine the kappa statistic.

The relationship between validity and reliability

Even if results are reproducible or reliable, they might not be valid. This will occur if a test consistently gives the wrong results (refer to figure 5-17). A test showing low reliability is illustrated in figure 5-18. Because the results cluster around the true value, they are valid for the majority of people within a population. Optimally tests should be reliable and valid (refer to figure 5-19 in Gordis).

3.7.1 Activity 3.4

Do the following activity and add it to your portfolio:

a) What is the difference between the validity and reliability of a screening test?
b) Name three factors that can contribute to variation between test results.
c) Complete the review questions at the end of chapter 5 in Gordis.

Learning unit 4

Ways of expressing prognosis

4.1 Introduction to learning unit 4

When talking about prognosis in relation to disease, it is essentially the prediction of the likely course and outcome of a disease and therefore includes the chance of recovery from an illness.

In this learning unit we will discuss the natural history of disease and various ways in which prognosis can be expressed. To work through the learning unit, refer to **chapter 6**, pages 116–137 in Gordis.

4.2 Learning outcomes

After completing this learning unit, you should be able to

- explain what each of these measures of prognosis is:
    - case fatality rate
    - person-years
    - five-year survival rate
    - observed survival, using a life table and the Kaplan-Meier method
    - median survival time
    - relative survival rate
- discuss the generalisability of survival data

4.3 Natural history and prognosis

Recommended reading: chapter 6, pages 116-135, in Gordis

The natural history of a disease is in essence the stages of a disease that occur in an individual. Refer to figure 6-2 in Gordis which outlines the natural progression of a disease in a patient. Detection and treatment of a disease at any stage can alter its natural progression, reducing the duration of illness and/or increasing the chance of recovery. The effects of treatment on prognosis can only be established if the natural history of the illness is known in the absence of treatment.

Prognosis is a prediction of the course of a disease based on the probability that a specific outcome will occur in the future. The forecasts are based on distinct groups of patients, although the outcome for each individual patient is quite different. For example, if there is a 10% chance of death and 90% chance of full recovery, some patients will die and some will live – both very different outcomes for the individual patient – although the majority of people will recover. Prognosis may be expressed in terms of deaths from the disease, survivors of the disease or in terms of the interval from diagnosis to the recurrence of the disease or from diagnosis to the time of functional impairment, disability or changes in the patient's quality of life. In this learning unit we will focus on prognosis in terms of death or survivors of disease.

When assessing prognosis, at what point do we begin to quantify survival time? Ideally it would be from the time of onset of the disease. The difficulty is that most people will only become aware that they are

ill once they have symptoms and not when they actually acquired the disease. To simplify the prognosis of a disease, duration of survival is counted from the time of diagnosis. However, variability occurs because the time of diagnosis is not necessarily the time at which the patient first became ill and patients will seek medical care at different points in the disease. Another factor is that patients who die before a diagnosis is made will not be included in the data gathered for disease prognosis. It is therefore important to remember that even if we say survivorship is measured from the time of diagnosis, the time frame may not be distinctly defined.

It is valuable to know the prognosis of a disease for several reasons:

- If the severity of the disease is known, then priorities for clinical services and public health programmes can be established.
- Patients often ask questions about prognosis. (If you have a disease you would want to know if you were going to get better, or if not, how much longer you would be expected to live.)
- If the normal disease process and chance of survival is known, then the effect of treatment can be compared to the expected outcome without treatment.
- The effectiveness of new and different types of treatments (as they become available) can be compared with one another and with older treatment regimes.

4.4 Measures of prognosis

There are a number of ways of expressing prognosis of death. A few common measures are

- case fatality rate
- person-years
- five-year survival rate
- observed survival rate
- median survival rate
- relative survival rate

4.4.1 Case fatality rate

Case fatality rate is a very useful way to express prognosis and we first discussed it in learning unit 2. Case fatality rate is defined as the number of people who die of a disease divided by the number of people who have the disease. There is no explicit declaration of time, but if the usual natural history of a disease is known, the expression “case fatality” refers to the time in which death might be expected to occur. For this reason it is typically used for a disease in which, if death is to occur, death happens soon after diagnosis, rather than for illnesses where death may occur many years after diagnosis.

4.4.2 Person-years

We spoke about person-time in learning unit 2. Please refer back to that section. A person-year is a measurement that combines the number of years that each person in a study was at risk of getting a particular condition, but did not get the condition and was observed. It is calculated by taking the sum of the number of years each person in an investigation was at risk of getting a specific condition.

It is most often used when determining incidence rates (number of deaths by the person-years over which a group is observed). Remember that when different individuals are observed for different lengths of time, we can calculate an incidence rate in which the denominator consists of the sum of the units of time that each individual was at risk and was observed. This is called person-time and can be expressed in terms of person-months or person-years of observation.

When working with person-years, we assume that each person-year is comparable to every other person-year. Of course, this may not be true, as there may be a time of greatest risk of dying. Refer to figures 6-3 to 6-7 on pages 118–119 of the textbook.

Despite this issue, person-years are useful in many situations such as randomised trials and cohort studies.

4.4.3 Five-year survival rate

The five-year survival rate is the percentage of patients who are still alive five years after diagnosis or treatment begins. It is often used in clinical medicine with regard to cancer treatments. The five-year interval was chosen simply because most deaths from cancer take place during the five-year period after diagnosis. Five-year survival rates are highly influenced by the time of diagnosis. For example, if someone is diagnosed earlier (owing to intensive screening programmes) they may appear to live longer (possibly more than five years), than if they had been diagnosed later, even though they may ultimately die at the same specific time. This is because five-year survival rate is from the time of diagnosis, not onset of disease. Refer to figures 6-8 and 6-9 in Gordis. This makes the assessment of the efficacy of screening programmes using five-year survival rates problematic.

The survival experience of the individuals in the population is also not taken into account; refer to figure 6-10 in Gordis.

4.4.4 Observed survival

Another approach to prognosis is to use the actual observed survival over time which expresses prognosis using life tables. A life table is essentially a statistical table which gives the survival data for a group of individuals. The probability of survival for any year after diagnosis can be calculated using the table. Refer to tables 6-1 to 6-9 and make sure you understand and can do the calculations to determine the probability of survival at any time after diagnosis.

The data from the tables and calculations includes

- the number of people alive and under observation at the beginning of each interval
- the number of people dying in each interval
- the number of people lost to follow-up each interval
- the conditional probability of survival of each interval
- cumulative probabilities of survival from the beginning of the study to the end of each interval

Note: an interval is a specific period such as a month or year. In the example in the textbook, the interval was a year.

The probability of survival can be graphically represented in what is called a survival curve; refer to figure 6-11 in Gordis.

4.4.5 The Kaplan-Meier method for calculating survival

The Kaplan-Meier method does not record death in predetermined intervals of time. Rather, the exact times of death are recorded. Refer to figure 6-12 on page 126, table 6-12 and figure 6-13 on page 128 in the textbook for an example of data analyses by the Kaplan-Meier method. Review the figures and table and make sure you understand the concepts and calculations involved in this simple example. Many studies on survival nowadays report data using the Kaplan-Meier method.

4.4.6 Life table usage

Life tables are used extensively in the analysis of clinical data and are the standard way in which survival between different groups is communicated and compared (refer to figures 6-15, 6-16 and 6-17 in the textbook for an example of a study dealing with survival data from life tables).

Two assumptions are typically made when using life tables:

1. There has been no change in the effectiveness of treatment or in survival during the time of the study.
2. The effectiveness of treatment or survivorship of people who are lost to follow-up is the same as the experience of those who are followed up.

A third factor that needs to be considered is that improvements in diagnosis may lead to changed prognosis. Advances in diagnosis should therefore be taken into account when calculating and interpreting survival data. See figures 6-18 and 6-19 in the textbook, which show the effects improved diagnosis may have on prognosis.

4.4.7 Median survival time

Median survival time is another approach to expressing prognosis and is defined as the length of time that half of the study population survives.

It has two advantages over mean survival time:

- It is less affected by extremes.
- To calculate median survival we only have to observe the deaths of half the study group; to calculate mean survival we have to observe the deaths of the entire group.

4.4.8 Relative survival rate

We want to compare the survival of any group of people with a disease to the survival we would expect in this age group even if they did not have the disease. This is especially relevant when applied to groups of older people who would not be expected to have 100% survival rates in the time frame studied. For this we use relative survival rate, which is defined as the observed survival in people with the disease divided by the expected survival if the disease were absent.

See table 6-13 and figure 6-20 on page 134 of the textbook.

4.5 Generalisability of survival data

Recommended reading: chapter 6, page 135, in Gordis

Not all data can be generalised to the widespread population. When hospital or clinic data is used, the people coming into the hospital or clinic are not representative of all patients in the community, as they generally have a number of conditions and therefore the data is biased. Refer to the study in figure 6-24 on page 127 of the textbook. Although the data may be biased, it may still be of value as long as this is kept in mind when the findings of these studies are being interpreted.

4.5.1 Activity 4.1

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

a) Why is the quantification of prognosis important?
b) Name four measures that are used to express prognosis.
c) Define the following:
   - case fatality rate
   - person-years
   - five-year survival rate
   - median survival time
   - relative survival rate
d) Why is case fatality ideally suited to cases that are short term and acute?
e) Describe one problem when using person-years.
f) Describe some of the limitations of five-year survival data.
g) Explain what a life table is and some of the assumptions that are made when using a life table.
h) Why should we not generalise all studies using survival data to different population groups?
i) How may improvements of diagnosis effect the calculation of prognosis even if there is no improvement in the success of treatment?
j) What is median survival time? What are the advantages of using median survival over mean survival?
k) Complete the review questions at the end of chapter 6 in the textbook.

IMPORTANT: Make sure you are familiar with all the calculations in this chapter.

Learning unit 5

Randomised trials and drug testing

5.1 Introduction to learning unit 5

The core intention in both public health and clinical practice is to change the natural history of a disease so that death or disability can be delayed and the health of the patient or population can be improved. To achieve this goal, the best available preventive or therapeutic measures must be selected. To help make this selection, randomised trials are carried out. They are used to evaluate the effectiveness and side effects of a new treatment. In this learning unit we will discuss possible study designs that can be used for evaluating new approaches to treatment and prevention.

To work through the learning unit, refer to **chapters 7 and 8**, pages 138–174 in Gordis.

5.2 Learning outcomes

After completing this learning unit, you should be able to

- explain the importance of randomised trials
- discuss the issues that have to be considered when designing randomised trials including
    - selection of subjects
    - allocation of subjects to treatment groups
    - data collection on subjects
    - crossover
    - factorial design
    - non-compliance
- explain the importance of sample size
- describe the phases involved in testing new drugs
- discuss the ethical concerns surrounding randomised trials and the importance of registering clinical trials to reduce publication bias

5.3 Randomised trials

Recommended reading: chapters 7 and 8, pages 139-163, in Gordis

The basic design of a randomised trial is as follows: we begin with a defined population that is randomised to receive either new or current treatment. Then we follow the subjects to see how many are improved in the new treatment group compared to the current treatment group (see figure 7-1 on page 139 of the textbook).

Some of the issues to be considered when designing a randomised trial:

- selection of subjects
- allocation of subjects to treatment groups
- data collection on subjects
- crossover

- factorial design
- non-compliance

Let's examine these in more detail.

5.3.1 Selection of subjects and allocation to treatment groups

It is important to clearly define, in writing, the criteria that will govern which subjects will be eligible for and selected in a study. They must be defined in such a way that if somebody were to read the selection criteria, they would select the same subjects that were chosen in the first place and therefore any study will in theory be reproducible.

5.3.1.1 Studies without comparison

In studies without comparison the subjects are not allocated to groups and no comparison is made with an untreated group or a group receiving some other type of treatment. This creates a predicament because we need comparison to say that the treatment had an effect on the subsequent outcome, otherwise we cannot describe what the true consequence of the treatment was and whether it had any effect at all.

5.3.1.2 Studies with comparison

It is important that studies have a control group for comparison and therefore after subjects have been selected, they need to be allocated to groups. There are a few methods that may be used:

- **Historical controls:** This involves comparing records of patients who were treated in the past before the new treatment became available to the outcomes of a disease that is being treated currently. Some problems with this include general changes in health over time unrelated to treatment which may affect outcome.
- **Simultaneous non-randomised controls:** A simultaneous control group is selected, but in a non-randomised manner. For example, people born on even days go in group A and those born on odd days go into group B. The most important thing is there must be no chance of selection bias.
- **Randomisation:** Randomisation during allocation to groups is the best way to reduce selection bias. This is because in random assignment the assignment of the next individual cannot be predicted and therefore modified. Tables of random numbers can be used to help assign people randomly into groups. This unpredictability makes it very hard for any subjective biases that the investigator might have to influence the selection of patients for one treatment group or the other. We hope that randomisation will increase the likelihood that the groups will be comparable in characteristics that may influence how people are affected by the condition, such as age or sex.

  It is important for both groups to be comparable based on variables such as age and sex, otherwise the results may be biased. The problem is that even if we match the groups on specific variables that we know about, there are still many variables that can affect prognosis that we do not know about and cannot measure. This is why randomisation is so important; it increases the likelihood that groups will be comparable in terms of variables that we recognise and variables that we are unaware of.

See figure 7-3 on page 145 of the textbook which illustrates how randomised and non-randomised selection of groups can influence the results of a study.

- **Stratified randomisation:** Although randomisation is likely to enhance the comparability of data, it does not assure it. If there is a known variable, such as age, that we believe could influence prognosis, an approach called stratified randomisation could be used.

  This method has two steps. In the first step the study population is separated into sections determined by each variable that may influence the outcome of treatment. The participants in each section are then randomly assigned to treatment groups. In this way you can ensure that the two groups are balanced and the data can be compared. Figure 7-4 on page 146 of the textbook shows how stratified randomisation works.

5.3.2 Data collection on subjects

It is very important that the data gathered for each of the study groups be of the same quality.

Data on the following attributes needs to be obtained:

- **Treatment**: It is essential for the treatment group to which the patient was assigned to be known and which therapy the patient actually received, for example whether they completed treatment.
- **Outcome**: The outcomes of treatment including both improvements/changes and side effects need to be recorded. It is very important that criteria be explicitly stated for all the outcomes to be measured and that all outcomes be measured comparably.
- **Prognostic profile at entry**: If there is a known risk factor for a bad outcome, we want to make sure that randomisation has resulted in groups that are comparable.

Masking or blinding may be conducted during a study. This occurs when the participants do not know which treatment they receive. This is performed to increase the objectivity of all the participants. If neither the participants nor the researchers know which treatment a subject receives, it is called double blinding. This objective is to stop researchers from treating the two groups differently.

5.3.3 Crossover

Crossover may also be utilised in clinical trials. In a planned crossover subjects begin the study on treatment A and later switch to treatment B and the subjects that started on treatment B move to treatment A (refer to figure 7-5 in Gordis). Patients can then serve as their own control, eliminating some effects owing to variation between individuals. Although this design is very appealing, certain factors must be considered:

- **Carryover**: This occurs when the effects of the initial treatment affect the observations for the second treatment. There has to be a washout period between treatments which ensures that the effects of the second agent are only due to that agent and not due to residual carryover from the first agent.
- **Psychological responses:** Individuals may feel and react differently when they start the first treatment versus how they behave when they start the second treatment. The researchers

need to ensure that any differences observed are indeed due to the agents being evaluated and are not due to differences in psychological responses or adherence to treatment.

Crossover studies are obviously not possible if the therapy cures the disease or if therapy results in a permanent change, for example surgery.

Unplanned crossover occurs when a subject that was assigned by randomisation to a particular group moves into the other group. For example, if group A was assigned to have surgery and group B was assigned to receive therapeutic drugs, a patient assigned to group A might refuse surgery and request to move to group B. Figure 7-6 in Gordis illustrates unplanned crossover. Unplanned crossover makes analysis of the data more complex, as the question then becomes which group do you assign the crossover participants to.

5.3.4 Factorial design

In factorial design the same study population is used for testing two different drugs, provided the drugs modes of action are distinct and that the expected outcomes for the two drugs are different. See figures 7-7, 7-8, 7-9, 7-10 and 7-11 on pages 151 and 152 of the textbook, which illustrate factorial design.

5.3.5 Non-compliance

Non-compliance occurs when patients do not act in accordance with the assigned treatment. Overt non-compliers refuse to participate and are referred to as dropouts. Covert non-compliers do not take the agent as requested and do not inform the researchers of their refusal to comply, thereby affecting the data.

5.3.6 Sample size

One of the most important decisions to make before a study is conducted is to decide how many subjects are needed for that study (sample size). Figures 8-1 to 8-3 illustrate how selection of the sample can influence results. A large enough sample is required to ensure that the results are accurate and indicative of what would be expected in a larger group of people with the same condition. If there are too few subjects in a study, at the end of the trial when the data is analysed, there may be no statistical difference in the data between two groups (treated and untreated), even though clinical significance exists. Therefore before a study is conducted, the sample size required for statistical significance needs to be estimated. This is done by specifying a number of factors (refer to table 8-3 in Gordis) which can then be used with sample size tables, formulae or computer programs to estimate the required sample size. Refer to the examples of sample size tables, tables 8-4 and 8-5, in Gordis.

5.3.7 Ways of expressing the results of randomised trials

There are several ways of conveying the results of randomised trials. One way is to calculate the efficacy of a drug or vaccine. The efficacy is calculated by subtracting the rate of developing disease in the people who received vaccine from the rate of developing disease in people who received the placebo and dividing this value by the rate of developing disease in the people who received the placebo. This will give us the extent to which the vaccine reduced the occurrence of the disease.

Other ways of analysing the data include

- calculating the ratio of the risks or relative risk in the two treatment groups
- comparing the survival curves of the groups
- estimating the number of patients who would need to be treated (NNT) to prevent one unfavourable outcome (such as death)

5.3.8 Generalising results

Clinical trials are conducted to find out what the effect of a particular drug would be on the general population and therefore the results are typically generalised beyond the study population itself. To determine whether it is possible to generalise the findings of a study, epidemiologists distinguish between the internal and external validity of a clinical trial. See figure 8-8 on page 162 of the textbook.

- **Internal validity** refers to the degree to which the results of a study are true for the studied population. It depends on the number of systematic errors in measurements and how well the study was conducted.
- **External validity** (generalisability) represents the degree to which the results of a study are relevant for populations other than the population that was initially tested.

A study that is not internally valid cannot be externally valid, but a study that is not externally valid may be internally valid and just not generalisable.

5.3.9 Activity 5.1

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

a) Why are randomised trials conducted?
b) What is the purpose of randomisation?
c) Name and discuss four issues that must be considered when designing a randomised trial.
d) What is stratified randomisation?
e) Discuss the two types of crossover.
f) What is a factorial design and under what circumstances can it be used?
g) Explain what non-compliance is. Why would covert non-compliance be a problem?
h) Why is sample size important? Describe how a sample size that is too small can result in invalid results,

5.3.10 Feedback on activity 5.1

It is important that you understand what a randomised trial is and why it is conducted.

When discussing four issues that must be considered when designing a randomised trial, did you discuss four of the following?

- selection of subjects
- allocation of subjects to treatment groups

- data collection on subjects
- crossover
- non-compliance

Stratified randomisation is illustrated in figure 7-4. Did you use this to help you answer the question?

When describing the two types of crossover, did you describe the difference between planned and unplanned crossover?

The selection of the correct sample size for a clinical trial is incredibly important. Make you understand how sample size can influence whether a clinical trial yields meaningful results.

5.4 Phases in testing new drugs in the United States

Recommended reading: chapter 8, pages 165-173, in Gordis

In South Africa the Medicines Control Council (MCC) is a statutory body that regulates the implementation of clinical trials and registration of drugs for use in South Africa. As the textbook deals with drug development in the United States, we will discuss drug development in that country. Keep in mind that although drug development procedures in South Africa will be similar, there may be a few differences. In the United States newly developed drugs have to pass through three phases of testing before they can be licensed for marketing.

- Phase I – small studies to assess the toxicity and pharmacological effectiveness of the drug (typically 20-80 patients)
- Phase II – testing of efficacy and relative safety in people (typically 100-200 participants)
- Phase III – large-scale randomised trials to further test efficacy and relative safety

Once licensed for marketing, it is important that phase IV monitoring studies be carried out to ensure that there are no adverse effects which may take years to develop or which occur so seldom that they are only detectable in very large populations.

5.4.1 Ethical considerations

There are many ethical issues involved in the use of clinical trials:

- Is it ethical to randomly withhold a drug from a patient if it may cure a serious or life-threatening disease?
- Is it ethical to use a placebo or to withhold a treatment that is known to help?
- Can true informed consent be obtained?
- In what situations should a trial be stopped earlier than originally planned? For example, should it be stopped if harmful or beneficial effects are noticed before the trial is finished?

5.4.2 Randomised trials for evaluating widely accepted interventions

Randomised controlled trials can be used for two main purposes:

- to assess new procedures or therapeutic agents before they are permitted and recommended for general use

- to appraise highly controversial interventions or practices that have been widely used or endorsed without having been sufficiently evaluated

Examples of each type of randomised trial are presented in the textbook. Read through the examples entitled “A Trial of Arthroscopic Knee Surgery for Osteoarthritis” and “Effect of Group Psychosocial Support on Survival of Patients with Metastatic Breast Cancer” on pages 169–172 in Gordis.

5.4.3 Registration of clinical trials

The International Committee of Medical Journal Editors have adopted a policy that states that all clinical trials of medical interventions must be registered in a public trials registry before any participants can enrol in the study. This is because not all results of clinical trials are published. This may result in publication bias when the results from all published clinical trials are reviewed.

For instance, if only the trials that show a beneficial result for a drug are published and the ones showing a non-beneficial or harmful result are never published, the reviewer will incorrectly assume that the drug is always beneficial. This form of selective reporting presents a significant risk to public health, and is the reason for the registration of all clinical trials.

5.4.4 Activity 5.2

Do the following activity and add it to your portfolio:

a) What are the principal phases involved in clinical trials in the United States?
b) What are some of the ethical concerns that need to be considered when conducting a clinical trial?
c) What is publication bias and why is it a problem?

Complete the review questions at the end of chapter 8 in Gordis.

Learning unit 6

Identifying the cause of a disease

6.1 Introduction to learning unit 6

In clinical medicine, being able to prevent and treat disease is important. When determining what measures can be used to prevent a disease, it is essential to understand what caused the disease (aetiology) in the first place. For example, if an environmental causative agent is involved, exposure to this agent can be eliminated or reduced. It is also valuable to know if an individual or group of people have an increased chance of suffering from a particular disease because of some shared characteristic risk factor. It is therefore essential to understand the different types of study designs that can be used to identify any risk factors that may be associated with a disease and to investigate its aetiology.

In this learning unit we will describe the cohort and case-control studies that can be used to determine the aetiology and risk factors for human diseases. To work through the learning unit, refer to **chapters 9 and 10**, pages 179–213 in Gordis.

6.2 Learning outcomes

After completing this learning unit, you should be able to

- describe what a cohort study is and when it is used
- explain the difference between a prospective and retrospective cohort study
- explain what biases may be present in cohort studies
- describe what a case-control study is
- discuss the selection of controls and the problems associated with the different selection methods
- describe what a case-control study based in a defined cohort is
- explain how a control group can be selected for in a case-control study
- describe case crossover design
- describe cross-sectional studies

6.3 Cohort studies

Recommended reading: chapter 9, pages 179-188, in Gordis

Cohort studies are suitable for estimating the risk of acquiring a disease and are used for determining whether there is an association between a factor or characteristic and the development of a disease. For example, if somebody is exposed to cigarette smoke, is there an increased risk of developing lung disease? Refer to figure 9-1 on page 179 of the textbook. If there are patterns of association between exposure to a particular factor and the development of disease, conclusions regarding a possible causal relationship can be made.

6.3.1 Design of a cohort study

A cohort study is useful for estimating the incidence of disease (or rate of death from disease) in two or more groups. First, the investigator selects two (or more) groups of individuals. One group has been exposed to the risk factor or disease and the other group has not been exposed to the risk factor (refer to figure 9-2 in Gordis). Both groups are then followed and the development of disease in the two different groups is compared. All the individuals of either group are observed until they either die, become a case, or the end of the study period is reached.

Refer to tables 9-1 and 9-2 in Gordis to see how incidence rates of the disease are calculated for each group. If the incidence in the exposed group is higher than the incidence in the non-exposed group, there is a positive association between the exposure and the disease.

Make sure you are able to calculate incidence rates of a disease if you are given the results of a cohort study.

6.3.2 Comparing cohort studies with randomised trials

Both cohort studies and randomised trials compare exposed groups with non-exposed groups (or a group with a certain exposure to a group with another exposure). The difference between the two designs is the presence (randomised trial) or absence (cohort study) of randomisation. Refer to figure 9-3 on page 169 of the textbook.

6.3.3 Selection of study populations

The fundamental characteristic of a cohort study is that a comparison is made between exposed and non-exposed people.

There are two ways to generate these groups:

- Select groups on the basis of whether or not they were exposed, so the study begins with exposed and non-exposed groups (refer to figure 9-4 in the textbook).
- Select a defined population before any of its members become exposed or before their exposures are identified, so the study begins with a defined population (refer to figure 9-5 in the textbook).

The drawback of the second study design is that it requires a long follow-up period.

6.3.4 Types of cohort studies

There are two types of cohort study designs:

- prospective cohort (concurrent cohort or longitudinal study)
- retrospective cohort (historical cohort or non-concurrent prospective cohort)

Refer to figures 9-6, 9-7 and 9-8 in Gordis.

Prospective cohort study

The investigator identifies the study population at the beginning of the study and monitors the subjects through time until the point at which the disease develops or does not develop. So essentially, the researcher starts the study at the time when it is realised that the cohort has been exposed to something or monitors when exposure occurs in a group of people. The investigator then determines the current status of the individuals, and selects only the individuals that have not yet developed disease and follows those individuals through time. Refer to figure 9-6 in Gordis.

As discussed earlier, the drawback of this approach is that because of follow-up, the study can take many years to complete. People are lost to follow-up and owing to improved methods of detecting exposure, there may be changes in classification of individuals.

Retrospective cohort study

The alternate approach to the prospective study design is the retrospective cohort study. In this type of design, the investigator uses historical data of all the exposed and non-exposed subjects and then assesses their case or non-case status (disease or no disease). So essentially, the study is started long after the exposure time and after the disease has established itself in the cohort being studied. The difference between the two designs is just a matter of time; in both designs, exposed and non-exposed populations are being compared. Refer to figures 9-7 and 9-8 in Gordis.

Drawbacks of this approach are selection bias and misclassification of individuals (depends on how reliable the records are) as historical data is being used.

Examples of cohort studies

Refer to the textbook and read the two examples of cohort studies entitled “The Framingham study” and “Incidence of breast cancer and progesterone deficiency”. Make a mental note of whether these studies were prospective or retrospective.

6.3.5 Cohort studies for investigating childhood health and disease

Cohort studies can be used to study childhood health and disease. Experiences and exposures during foetal life may have long-lasting effects on the individuals; these effects can be investigated using long-term cohort studies. An example is a study of the effects of exposure to radiation in utero on the development of cancer later in life.

Studies dealing with large groups of children can provide a great deal of information about exposure and development of certain conditions. However, because we are dealing with a large cohort who needs to be studied over a long time, there are a number of important factors that need to be considered:

- At what point should the individuals in the cohort be identified? (Refer to figures 9-10 and 9-11 in Gordis.)
- Should the cohort be drawn from one place or from a few places, or should a national sample be drawn?
- For how long should the cohort be followed?
- What hypothesis and how many hypotheses should be tested in the cohort?

6.3.6 Potential biases in cohort studies

When conducting cohort studies, there are certain biases that should either be avoided or taken into account. Some potential biases include

- bias in assessment of the outcome
- information bias
- biases from non-response and losses to follow-up
- analytic bias

6.3.7 When is a cohort study warranted?

A cohort study is generally justified when there is enough evidence to suggest an association of a disease with a certain exposure or exposures. It can then be used to confirm this association. Refer to figure 9-12 in Gordis.

Even though the cohort study design is very useful when investigating the causes of disease, it is impractical, for instance, if there is not enough strong evidence of the role of a specific risk factor in the aetiology of a disease or if the disease occurs at a very low rate. Large studies are expensive and if there is not enough evidence to warrant a study, no organisation will want to fund it.

Cohort studies require the populations being studied to be followed until the development of disease and may require follow-up over a long time. It is therefore less challenging to conduct a study if the interval between exposure and development of the disease is short.

6.3.8 Activity 6.1

Do the following activity and add it to your portfolio.

Remember, the activities serve as part of your preparation for the exam!

a) Explain how a cohort study is conducted.
b) Describe the difference between cohort studies and randomised trials.
c) Name and describe the two types of cohort study designs.
d) What are some of the potential biases that can result when conducting a cohort study? Discuss each type of bias.
e) Answer the review questions at the end of chapter 9 in Gordis.

6.3.9 Feedback on activity 6.1

When describing the two types of cohort study designs, you should have described a prospective cohort study and a retrospective cohort study.

You should have discussed bias in assessment of the outcome, information bias, biases from non-response and losses to follow-up and analytic bias

6.4 Case-control studies and other study designs

Recommended reading: chapter 10, pages 189-212, in Gordis

In the previous learning unit we discussed cohort study design as a means of investigating the aetiology of disease. In this learning unit you will learn about other study designs that can be used to study the aetiology of disease, including

- case-control studies
- case crossover design
- cross-sectional studies

6.4.1 Case-control study

In all of these designs comparison is used and is a necessary component of epidemiological investigation. Without comparison (a control group), it would be impossible to determine whether a certain exposure is related to the risk of developing some condition. During a case-control study, people with the disease, called cases, are compared to people without the disease, called controls. In these two groups the level of exposure to some agent or risk factor is then compared. This is in contrast to the cohort study, which begins with a group of exposed people and then compares them to a non-exposed group.

Therefore, what differentiates the two designs is whether the study begins with diseased or non-diseased people (case-control study) or with exposed and non-exposed people (cohort study).

6.4.1.1 Design of a case-control study

The design of a case-control study is illustrated in figure 10-1. A case-control study consists of the following steps:

1. Identify a group of individuals with the disease (cases) and a group of individuals without the disease (controls).
2. Determine what proportion of the cases were exposed to a particular factor and what proportion were not.
3. Determine what proportion of the controls were exposed to the same factor and what proportion were not.

If the answer to whether exposure occurred is either yes or no, then there are four groups of people:

- cases that were exposed
- cases that were not exposed
- controls that were exposed
- controls that were not exposed

The proportion of cases exposed and the proportion of controls exposed can then be calculated. This is illustrated in table 10-1 in Gordis.

At this point it is important to note that data from a case-control study cannot be used to estimate the

prevalence of the disease. This is because the investigator determines the number of controls per case. Therefore the proportion of the study population that consists of cases is determined by the investigator and does not reflect the true prevalence of the disease in the population.

Refer to tables 10-3 and 10-4 in the textbook which show the results of two case-control studies. Also note that in the second study dealing with the link between lung cancer and smoking, the exposure is not just classified as exposed or not exposed, but rather the level (number of cigarettes a day) is also accounted for.

6.4.1.2 Selection of cases and controls

Cases can be selected from a number of sources such as hospital, clinic or private physician patients. When selecting cases, care must be taken to ensure that the risk factors that are identified are generalisable to all patients with the disease and the criteria for selecting those cases must be clearly stated and specified in writing.

Selection of cases

When designing a case-control study, an investigator can use incident cases of a disease (newly diagnosed cases) or prevalent cases of the disease (people who have had the disease for some time). When prevalent cases are used, the researcher does not have to wait for new cases to be diagnosed and larger numbers of cases are readily available for study. In spite of this, the use of incident cases is preferred because the risk factors identified in prevalent cases may be associated more with survival with the disease than with the development of the disease.

Regardless of which cases are chosen, it is important to allow for any selection biases that may have been introduced when selecting cases.

Selection of controls

One of the most challenging aspects of designing a case-control study is selecting an appropriate control group. The way in which the controls are selected will determine whether the conclusions about the association between an exposure and a disease are valid. Refer to the example study that tested the hypothesis that tuberculosis protects people against cancer on page 192 in Gordis conducted by Raymond Pearl. This study clearly shows how selecting the incorrect control group can result in incorrect conclusions.

Sources of controls

Controls can be selected from

- non-hospitalised people
- hospitalised patients (with diseases other than the disease for which the cases were admitted)

The non-hospitalised controls can be selected from several sources in the community such as

- school rosters
- insurance company lists
- a control for each case (e.g. someone from the neighbourhood in which the case lives -

neighbourhood control)
- using a best friend, spouse or sibling as control

The difficulty when selecting patients from hospitals as controls is that it is generally not possible to characterise the reference population from which hospitalised cases come and hospital patients may differ from the people in the general community. The investigator will also have to decide whether to use all other patients admitted to the hospital as controls or whether to select a specific group of people as controls. The concern with selecting a specific diagnostic group is that these controls are unlikely to be representative of the general reference population.

6.4.1.3 Problems in control selection

Great care must be taken when selecting controls because the controls might have a very high or a very low level of exposure that might not be representative of the level in the population of the study.

When a difference in exposure is observed between cases and controls, it must be considered that the level of exposure observed in the controls might not be the level of exposure expected in the study population. Refer to the study on the relationship between pancreatic cancer and coffee consumption on page 196 in Gordis.

6.4.1.4 Problems of memory

When exposure data is collected, the quality of the data collected will depend on the ability of the cases and controls to accurately recall any exposures. During case-control studies, data related to exposure is typically collected from the subjects in interviews. Unfortunately the information obtained during these interviews may not always be very accurate. This is because

- most human beings are limited in their ability to remember information
- the subjects may not have the information being requested

If subjects do not accurately know their exposure status, misclassification of exposure status will result. Some subjects that were actually exposed will be wrongly classified as unexposed and unexposed individuals may be classified as exposed. This may result in an underestimate of the risk of disease associated with exposure.

Recall bias, also called rumination bias, is also a problem. It may occur when the information provided on exposure is different between the cases and controls because the cases have spent more time thinking on what could have resulted in the disease and are therefore more likely to remember potential exposures. Refer to table 10-10 in Gordis.

6.4.1.5 Matching cases and controls

For the data generated to be valid, it is important to ensure that the cases and controls have similar characteristics such as age, sex and socioeconomic status. Matching is a method used to ensure that the cases and controls are similar in certain characteristics or exposures that might influence a factor that is being studied. There are two types of matching:

- **Group matching:** The cases are selected first, and then the controls are selected so that the percentage of the controls with a certain characteristic is identical to the percentage of cases with the same characteristic. For instance, if 30% of the cases are female, then 30% of the controls will be female.
- **Individual matching:** A control is selected with similar characteristics for each case. Each case is therefore individually matched to a control based on variables that might be of concern. Even though epidemiological research often involves matching, there are some problems associated with it. If too many variables are matched, it may be impossible to find an appropriate control. Unplanned matching that may bias the control group can also occur.

6.4.1.6 Use of multiple controls

During a case-control study it is up to the investigator to decide how many controls will be used per case. Multiple controls can be used for each case and they may be either controls of the same type or controls of different types.

Multiple controls can be useful to explore alternative hypotheses and for analysing possible biases, such as recall bias. Refer to the case-control study of brain tumours in children on pages 200–202 of the textbook.

6.4.1.7 When is a case-control study warranted?

Case-control studies are typically conducted to find the cause of a disease and can be used to investigate the potential roles of a number of exposures or factors that may be associated with a disease. If a case-control study indicates an association between a specific exposure and the disease, then a cohort study can be conducted to study the relationship in further detail. Case-control studies are generally less expensive than cohort studies and can be carried out more quickly. They also offer a more efficient design for investigating rare diseases.

6.4.2 Case-control studies based in a defined cohort

Recommended reading: chapter 10, pages 203-206, in Gordis

Some elements of both a cohort and case-control study can be combined into a single study, where a case-control study is initiated within a cohort study. This is shown in figure 10-8 in Gordis. In this type of design, the cases come from the same cohort as the controls to which they are compared.

Cohort-based case-control studies can be divided into two types:

- **Nested case-control studies:** As each case of the disease develops, a control is selected from the individuals who are at risk for the disease at the time. Therefore, cases and controls are matched on calendar time and length of follow-up (refer to figure 10-9 in Gordis).
- **Case cohort studies:** Unlike nested case-control studies, the controls are not individually matched to each case, but are chosen randomly from the defined cohort with whom the study began (refer to figure 10-10 in Gordis).

Advantages of embedding a case-control study in a defined cohort

- The problem of possible recall bias is eliminated.
- Abnormalities in biological characteristics are more likely to be an indication of risk factors rather than a manifestation of subclinical disease.
- It is often more economical to conduct.
- Cases and controls are derived from the same original cohort, so there is likely to be greater comparability between the cases and controls than is normally found in a traditional case-control study.

6.4.3 Activity 6.2

Do the following activity and add it to your portfolio:

a) Explain how you would design a case-control study.
b) Can the data from a case-control study be used to estimate the prevalence of a disease? Why?
c) How would you select cases and controls and what are some of the problems that must be kept in mind when selecting them?
d) Discuss matching and some of its drawbacks.
e) When is a case-control study warranted?
f) What is a case-control study based in a defined cohort?

6.4.4 Case crossover design

Recommended reading: chapter 10, pages 206-210, in Gordis

This type of study is used mainly to examine the effects of transient exposures (such as air pollution) on the risk of acute illness (such as asthma). Each case acts as its own control just at a different time under different conditions. Therefore all subjects are cases and the chance of experiencing an acute-onset disease is assessed under two different conditions or exposures. It is assumed that if an exposure or event is linked to disease onset, then the exposure or event should happen more often just before disease onset than at times when there is no acute disease. Refer to figure 10-11 in Gordis.

Essentially this type of study examines whether a risk factor was present or absent immediately before the outcome or disease was experienced and whether the same risk factor was present when the outcome was not experienced. This will clarify whether there is a link between the risk factor and the outcome.

6.4.5 Cross-sectional studies

Recommended reading: chapter 10, pages 210-212, in Gordis

In this study design the disease and exposure status are measured simultaneously for each subject, providing a snapshot of the frequency and characteristics of a disease in a study population at a particular point in time. See figures 10-13 and 10-14 on page 211 of the textbook.

This study design provides weak evidence of causal association between exposure and outcome because the exposure may not have occurred before the onset of the disease. Cohort and case-control studies need to be conducted to establish the causal association between exposure and disease or outcome.

6.4.6 Activity 6.3

Do the following activity and add it to your portfolio:

a) Name and discuss some of the other study designs that can be used for studying the aetiology of a disease.

b) Complete the review questions at the end of chapter 10 in Gordis.

Learning unit 7

Causal inferences: bias, confounding and interaction

7.1 Introduction to learning unit 7

In the previous units, we discussed the designs of epidemiological studies that are used to determine whether there is an association between an exposure and a disease. If there is an association between the exposure and the disease, the next step will be to determine whether the observed association reflects a causal relationship. It is important for an epidemiologist to realise that an association that is observed during a study may be causal, non-causal or owing to chance, confounding or bias.

In this learning unit we will focus on three very important issues that must be considered when deriving causal inferences:

- bias
- confounding
- interaction

To work through the learning unit, refer to **chapter 15**, pages 262–277 in Gordis.

7.2 Learning outcome

After completing this learning unit, you should be able to discuss and explain how the following factors can influence epidemiological studies:

- bias
- confounding
- interaction

7.3 Bias

Recommended reading: chapter 15, pages 262-266, in Gordis

Bias occurs as a result of an error in the design, organisation or analysis of a study and may result in an incorrect assessment of an exposure's influence on the risk of developing a specific disease. Therefore efforts should be made to identify, reduce or eliminate bias. Selection bias and information bias are two common types of bias that are encountered in epidemiological studies.

7.3.1 Selection bias

If the cases and controls (or exposed and non-exposed individuals) that were selected showed some noticeable link, between exposure and disease for example, but there was no real association, this may have occurred as a result of selection bias.

One form of selection bias is **non-respondent bias** and it occurs when the participants that respond to a survey respond in a way that is different from the people who do not respond. For example, if a questionnaire about lung issues and smoking is sent out, smokers may be less likely to respond if they are currently having problems breathing as they may not wish to acknowledge that smoking is bad for them. The results of the study will then be biased and will not reflect the characteristics of the general population because the portion of smokers with current lung issues is not taken into account when they never responded to the questionnaire.

Another form of selection bias is **exclusion bias** and it occurs when the investigator uses dissimilar eligibility criteria to select the cases and controls. This may be because the investigator feels that certain clinical conditions or characteristics in the control group should result in their exclusion. However, if the same exclusion is not applied to the cases the data may become biased. Refer to the study where women with breast cancer were compared to women without breast cancer in terms of use of reserpine on page 263 in Gordis.

7.3.2 Information bias

Information bias can arise when the way information is gathered about the participants in a study is not complete or accurate enough. This may result in a portion of the information related to exposures and/or disease outcome being wrong.

Misclassification bias results when participants are misclassified, for instance when people who have the disease (cases) are misclassified as controls (they may not yet have been diagnosed), or controls are misclassified as cases. Exposure status may also be inaccurate if a person is not aware that they have been exposed or feels that they were exposed but in actual fact they were not.

There are two types of misclassification:

- differential misclassification (the rate of misclassification varies in the separate study groups)
- non-differential misclassification (problem with the accuracy of data collection from both separate study groups, cases and controls or exposed and unexposed)

Some forms and sources of information bias include

- bias in abstracting records
- bias in interviewing
- bias from surrogate interviews
- surveillance bias
- recall bias
- reporting bias

As bias can have a significant effect on the findings of a study, it is important to reduce or eliminate it. If this is not done, it should at least be recognised and accounted for.

7.3.3 Activity 7.1

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

a) Define bias and explain the two main types of bias we encounter in epidemiological studies.
b) Explain what non-respondent bias, exclusion bias and misclassification bias are and how they may impact on the findings of an epidemiological study.
c) Briefly describe the following types and sources of information bias:
    - bias in abstracting records
    - bias in interviewing
    - bias from surrogate interviews
    - surveillance bias
    - recall bias
    - reporting bias

7.4 Confounding

Recommended reading: chapter 15, pages 266-270, in Gordis

During an epidemiological study when a true association is observed, it might be tempting to say, for example, that a specific factor caused a specific disease when in actual fact it did not. If this happens, then the observed relationship is not causal and may rather be as a result of confounding. Confounding is an apparent association between disease and exposure caused by a third factor not taken into consideration.

Say we find a link between a particular factor, called factor A, and a disease. We may then assume that factor A causes the disease. But factor X which is associated with factor A (but is not a result of factor A) may also be a risk factor for the disease. Factor X may then be the real factor that led to the development of disease and not factor A, but because only factor A was studied, the results seem to indicate that it is associated with the development of the disease. Refer to the examples displaying confounding figures 15-1 and 15-2 and tables 15-4 to 15-7 in the textbook on pages 267–268.

In short, a confounder is a third factor that must be considered when deciding whether an association is causal.

The problem of confounding can be dealt with either by

- designing and carrying out a study where the cases are matched to the controls for the factor suspected to be a confounder
- analysis of the data by stratification or adjustment

Refer to tables 15-9 and 15-10 and figures 15-3, 15-4 and 15-5 which deal with stratification of data.

Even though a confounder is generally considered to be a problem, finding a confounded relationship can also be very useful as it can still identify people who are at high risk for the disease. It is also important to understand that confounding is not an error in the study; it is a true occurrence that must be taken into account when interpreting the results of a study so the conclusions of the study are not biased.

7.4.1 Activity 7.2

Do the following activity and add it to your portfolio:

a) In your own words describe what confounding is.
b) Give an example from the textbook of a study where confounding occurred.
c) Explain how the problem of confounding can be resolved.

7.5 Interaction

Recommended reading: chapter 15, pages 270-276, in Gordis

In reality, very few exposures cause disease entirely by themselves. More often, multiple causal factors interact to cause a disease. Interaction occurs when two or more risk factors modify the incidence of disease and the joint effect of the two causal factors differs from what would be expected when adding their effect when they are acting independently.

A positive interaction (synergism) has occurred when the combined effect of the two factors is greater than what we would expect. A negative interaction (antagonism) has occurred when the combined effect of the two factors is less than what we anticipate.

The first step when looking for an interaction is to establish if there is confounding. If not, then is there interaction (refer to figure 15-6)? If the association is equally strong in different strata that were determined according to a third variable, then there is no interaction. If there is interaction, then to calculate the incidence after exposure to both factors, one of the following models can be used:

- additive model
- multiplicative model

Refer to tables 15-11 to 15-22 and the associated text on pages 271–276 of the textbook on interaction. The choice of which model to use to explain the interaction usually depends on the biology of the disease.

7.5.1 Activity 7.3

Do the following activity and add it to your portfolio:

a) Describe the role of interaction in the risk of acquiring a particular disease.
b) What models can be used to calculate the incidence after exposure to two different factors?
c) Now, complete the review questions at the end of chapter 15 in Gordis.

Learning unit 8

Roles of genetic and environmental factors in disease causation

8.1 Introduction to learning unit 8

You already know that environmental factors play a role in the development of certain diseases, but you must also take into account the part played by genetic factors. Disease does not develop in everybody exposed to an environmental risk factor. This is because human beings differ from one another and some people will be more susceptible to a disease than others because of their genetic makeup.

In this learning unit we will investigate the aetiology of disease by examining the role played by genetic factors and environmental factors and how the interaction of these two factors can either increase or decrease the risk of disease. To work through the learning unit, refer to **chapter 16**, pages 279–304 in Gordis.

8.2 Learning outcomes

After completing this learning unit, you should be able to

- describe what genetic markers are and how they can used by epidemiologists
- describe how studies examining twins, first-degree relatives and adopted children can be used to determine the relative contributions of genetic and environmental factors to disease causation
- discuss how international studies and migration can aid in understanding the genetic contribution to the aetiology of disease

What follows are some of the approaches that epidemiologists use to determine the relative contributions of genetic and environmental factors to disease causation.

8.3 Association with known genetic diseases

Recommended reading: chapter 16, pages 279-281, in Gordis

If we wish to investigate whether a particular disease has a strong genetic component, we can ask whether the disease is associated with other diseases or conditions that are known to have strong genetic components. Table 16-1 shows a number of diseases that are associated with diseases of known genetic origin. If there is an association between a condition of interest and a disease that has a known genetic cause, it does not confirm that the disease is genetically determined. However, it does indicate that some aspects of the aetiology of the disease or some instances of the disease may be due to genetic factors.

When dealing with a disease that occurs in hereditary and non-hereditary forms, we can try and identify the genes responsible for the hereditary form in the hope that they will provide a clue to the role of genetic factors in the non-hereditary cases.

Read the example on pages 280-281 in Gordis which discusses breast cancer that occurs in hereditary and non-hereditary forms in the section “Association with known genetic diseases”.

8.3.1 Activity 8.1

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

a) Why does disease not develop in everybody exposed to a certain environmental risk factor?
b) Why is it helpful to determine whether a disease (with a genetic component) is associated with a known genetic disease? Give an example.

8.4 Use of genetic markers

Recommended reading: chapter 16, pages 282-285, in Gordis

A genetic marker is a gene, DNA sequence or gene product that is associated with a specific condition or risk of obtaining a specific condition that can be tested for in a laboratory. Advances in molecular biology have resulted in the development of methods that can be used to test for and evaluate genetic markers and thus give information on the relationship between a disease and its genetic cause.

DNA microarray technology involves examining the levels of expression of large numbers of genes simultaneously and not just what genes are present in an individual. This is advantageous because someone may have a particular gene but it may not be being expressed or may be being overexpressed, and these differing levels of expression may contribute to disease.

HLA (human leukocyte antigen) types are genetically determined and have been found to be linked to certain diseases (refer to table 16-2). If disease associations are known for certain HLA types, then these genetic markers can be used to identify population subsets that have an increased risk of developing disease.

8.4.1 Activity 8.2

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

a) What is a genetic marker?
b) How can genetic markers be used to determine risk of disease?

8.5 Age of onset of disease

Recommended reading: chapter 16, pages 285-287, in Gordis

When a disease occurs in both genetic and non-genetic forms, the genetic form tends to develop in patients at a much earlier age than the non-genetic form (refer to figure 16-2 which illustrates the age of onset of retinoblastoma in familial versus non-familial cases). This is explained by the fact that non-genetic diseases require a build-up of environmental exposures over time and therefore disease takes longer to develop than in cases where the disease is genetic in origin. Refer to figures 16-3, 16-4 and 16-5 on pages 286–287 of the textbook which illustrate how the age of onset of disease is affected by genetic contributions.

8.5.1 Activity 8.3

Do the following activity and add it to your portfolio:

When disease occurs in both genetic and non-genetic forms, the genetic form develops at a much earlier age than the non-genetic form. Why is this so?

8.6 Family studies

Recommended reading: chapter 16, pages 287-296, in Gordis

Certain diseases run in certain families and may be as a result of genetic or environmental factors. It makes sense that specific genetic traits that make people susceptible to disease will be present in different family members, but it is also true that certain environmental exposures will be shared by family members. A number of different types of family studies can be conducted to obtain information on the environmental and genetic contributions to disease.

8.6.1 Risk of disease in first-degree relatives

A person with a certain disease, who has a large number of first-degree relatives with the same disease, suggests a genetic component to the disease although it does not prove a genetic link. Family pedigrees shed light on how a disease is passed from generation to generation and can be used to estimate the influence of genetic traits on the cause of the disease (refer to figure 16-6). If a husband and wife both have the disease (spousal concordance), then environmental factors are most likely the cause.

Applying molecular biological methods to family studies

If there is an increased risk of a disease in a family, molecular biology and epidemiological techniques can be utilised to determine whether there is a particular gene associated with the higher risk of disease. This is accomplished by doing segregation and linkage analysis.

8.6.2 Twin studies

Studying twins offers a unique opportunity to ascertain what contributions the genetic and environmental factors make to the development of disease. There are two types of twins, monozygotic (identical) and dizygotic (fraternal). Monozygotic twins come from the same fertilised ovum and are thus genetically identical. Dizygotic twins come from two separate ova and are the same as ordinary siblings that happen to develop in the uterus at the same time.

Regarding the occurrence of disease in monozygotic twins, both may have the disease or neither may have it (concordant for the disease) or only one twin may have the disease while the other twin does not (discordant for the disease). Concordance may occur as a result of genetic factors or environmental factors but if the twins are discordant for the disease, the disease is most likely environmental in origin. If a disease is genetic, we would expect less concordance for dizygotic twins (or siblings) than for monozygotic twins. Figure 16-10 and the associated text describe how the rates of concordance and discordance in twins are calculated. Refer also to tables 16-4, 16-5 and 16-6 which give details on

various studies on concordance data in monozygotic and dizygotic twin pairs.

Publication bias is a big problem when working with twin concordance data, as it is more likely that a researcher will publish data when both twins have a rare disease than if they observe only one twin with a disease.

Look at the examples in the textbook in figures 16-11 and 16-12 and tables 16-7 and 16-8 dealing with various twin studies.

8.6.3 Adoption studies

Monozygotic twins (who share 100% the same genes) normally grow up in the same environment. This makes it difficult to determine what contribution the environment made and what contribution genetic factors made. One way to uncover whether a disease is primarily genetic or environmental in origin involves studying children who were adopted by different families.

For instance, if we wanted to determine whether schizophrenia is primarily genetic or environmental in origin, we can conduct a study by

- examining offspring of non-schizophrenic biological parents who are adopted by schizophrenic parents
- examining offspring of schizophrenic parents who are adopted by non-schizophrenic parents
- examining offspring of non-schizophrenic biological parents who are adopted by non-schizophrenic biological parents

Refer to pages 295 - 296 and table 16-10 in Gordis covering the study of the inheritance of schizophrenia and adoptees.

8.6.4 Activity 8.4

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

a) When a disease aggregates in a family, what does it tell us about the contributions made by genetic and environmental factors?
b) What are segregation analysis and linkage analysis? When are these tests used and what information do they yield?
c) Describe how twin studies can be used to examine whether a disease is primarily genetic or environmental in origin.
d) What factors need to be taken into account when conducting adoption studies? Remember to discuss age at adoption and contact with biological parents.

8.7 Time trends in disease incidence

Recommended reading: chapter 16, page 296, in Gordis

If the number of cases of a disease increases or decreases significantly over time, then it can be

assumed that environmental factors are most likely the cause of the disease. This is because the overall genetic characteristics of the human population do not typically change over short periods. Refer to figure 16-14 on page 297 of the textbook.

8.8 International studies

Recommended reading: chapter 16, pages 297-299, in Gordis

Disease rates differ significantly across the world. Refer to figures 16-15 and 16-16 on page 297 of Gordis. In these figures you can see that Japan has one of the lowest rates of breast cancer in women in the world but the highest rate of stomach cancer in men. Are these differences due to variations in

- access to medical care?
- quality of medical care?
- differing record keeping or diagnosis in the different countries?

If the data is real, and other studies suggest it is, then are the differences in disease incidence due to genetic or environmental factors? To determine what contribution is made by genetic factors and what contribution is made by environmental factors, migrants can be studied in a way that is similar to how adoptees are studied. Read the section on migrant studies on pages 297-299 of the textbook.

8.9 Interaction of genetic and environmental factors

Recommended reading: chapter 16, pages 299-302, in Gordis

Certain diseases are largely environmental, whereas others are largely genetic. In some diseases, however, both environmental and genetic factors play a significant role. It is, therefore, important to understand the interaction between the two types of factors when studying disease causation. In diseases that are mostly caused by environmental factors, heritability of the disease is low. However, when the environmental causes are successfully identified and removed, there may still be cases in which genetic factors play the major role. Refer to figures 16-17 and 16-18 which show how environmental and genetic factors may contribute to a particular condition.

Recently, molecular studies have helped epidemiologists understand the contributions of environmental and genetic factors to certain conditions. This is accomplished by confirming that a disease is not only caused by environmental factors when specific genetic markers are associated with an increased incidence of disease and that certain environmental agents can have an effect on certain genes.

8.9.1 Activity 8.5

Do the following activity and add it to your portfolio.

Remember, this could serve as part of your summary to use in preparing for the exam!

a) If a time trend is observed in disease, does it implicate environmental or genetic factors in the disease causation? Explain.
b) When there are international differences in the risk of a disease, how are migrant studies employed to determine if the difference is due primarily to genetic factors or environmental factors?

c) Discuss some problems with migrant studies (refer to table 16-14 in Gordis).

d) Explain the interaction of environmental and genetic factors in disease causation.

e) Complete the review questions at the end of chapter 16 in Gordis.

8.10 Conclusion

In instances where both genetic and environmental factors play a role in the development of disease, determining the contribution of each factor to the cause of the disease is a challenging task. This is because each person has a range of physical, genetic and environment susceptibilities which all interact to determine their overall susceptibility to disease. It is, therefore, important to improve our understanding of genetic changes and individual susceptibilities within a population.

Forum 1: Student Lounge

Use this forum to discuss general matters among yourselves

Discussion 1: Introduce yourself

Use this space to get to know your fellow classmates.

Tell one another about your current work situation, professional background and something personal (± 250 words).

Discussion 2: Fellow student contact details

Use this space to share your contact details with your classmates and to form study groups.

Forum 2: Learning units

Use this forum to discuss specific topics you are trying to understand in the course

Forum 3: South African disease outbreaks (course discussion topic)

Use this forum to discuss your findings from the question posed in activity 1.4

Announcement 1: Welcome and getting started

Dear Student

Welcome to Epidemiology. We are happy to announce that this module will be presented online in 2015. You should have received a Getting Started letter in the mail explaining what we expect of you as an online student.

But for now, we would like you to first go to the **Discussion Forums** link on the left side of your screen and access Forum 1: Student Lounge. In Discussion 1 we want you to participate in your first online activity where you need to introduce yourself to your fellow students. Please participate in this discussion during February 2015.

We are looking forward to meeting you online!

Your lecturers